

PERCEPTION predicts patient response and resistance to treatment using single-cell transcriptomics of their tumors

Received: 20 June 2023

Accepted: 8 March 2024

Published online: 18 April 2024

 Check for updates

Sanju Sinha  ,
Rahulsimham Vegezna^{1,17}, Sumit Mukherjee ,
Ashwin V. Kammula  , Saugato Rahman Dhruba , Wei Wu  ,
D. Lucas Kerr , Nishanth Ulhas Nair , Matthew G. Jones^{4,5,6,7}, Nir Yosef^{4,5},
Oleg V. Stroganov⁸, Ivan Grishagin^{8,9}, Kenneth D. Aldape  ,
Collin M. Blakely  , Peng Jiang , Craig J. Thomas  , Cyril H. Benes  ,
Trever G. Bivona  , Alejandro A. Schäffer  & Eytan Ruppin  

Tailoring optimal treatment for individual cancer patients remains a significant challenge. To address this issue, we developed PERCEPTION (PERsonalized Single-Cell Expression-Based Planning for Treatments In ONcology), a precision oncology computational pipeline. Our approach uses publicly available matched bulk and single-cell (sc) expression profiles from large-scale cell-line drug screens. These profiles help build treatment response models based on patients' sc-tumor transcriptomics. PERCEPTION demonstrates success in predicting responses to targeted therapies in cultured and patient-tumor-derived primary cells, as well as in two clinical trials for multiple myeloma and breast cancer. It also captures the resistance development in patients with lung cancer treated with tyrosine kinase inhibitors. PERCEPTION outperforms published state-of-the-art sc-based and bulk-based predictors in all clinical cohorts. PERCEPTION is accessible at <https://github.com/ruppinlab/PERCEPTION>. Our work, showcasing patient stratification using sc-expression profiles of their tumors, will encourage the adoption of sc-omics profiling in clinical settings, enhancing precision oncology tools based on sc-omics.

In recent years, precision oncology has made important strides in advancing treatment for patients with cancer, as described in several reviews^{1–6}. Much of the focus in the field has been on efforts to use FDA-approved sequencing assays to identify ‘actionable’ mutations in cancer driver genes, to match patients to treatments¹. These efforts have been further boosted by the progress made in DNA-based liquid biopsies, which can further help guide and monitor treatment^{7–9}. However, a large fraction of patients with cancer still do not benefit from such targeted therapies, and therefore efforts are needed to find ways to analyze other molecular omics data types to benefit more patients. Addressing this challenge, recent studies have begun to explore the

benefit of collecting and analyzing bulk tumor transcriptomics data to guide cancer treatment^{10–17}. Expression-based studies have demonstrated the potential to complement DNA sequencing approaches in increasing the benefit of omics-guided precision treatments to patients.

One key limitation of current genomic and transcriptomic treatment approaches is that they are mostly based on bulk tumor data. Tumors are typically heterogeneous and composed of numerous clones, making treatments targeting multiple clones more likely to diminish the likelihood of resistance emerging owing to clonal selection, and hence potentially enhancing the overall response of

the patients¹⁸. Intra-tumor heterogeneity has been driving two major developments in recent years: the search for effective treatment combinations and the advent of single-cell (sc)-profiling of the tumor and its microenvironment.

Large-scale combinatorial pharmacological screens have been performed in patient-derived primary cells (PDCs), xenografts and organoids, and have already given rise to numerous combination treatment candidates^{19–21}. Concomitantly, the characterization of the tumor microenvironment via sc-omics has led to important insights regarding the complex network of tumor–microenvironment interactions involving both stromal and immune cell types¹⁸. It also offers a promising way to learn and predict drug response at an sc resolution. The latter, if successful, could guide the design of drug treatments that target multiple tumor clones disjointly^{22–24} and help us understand the ensuing resistance to better overcome it. However, building such predictors of drug response at an sc resolution is currently challenging owing to the paucity of large-scale preclinical or clinical training datasets. Previous efforts, including a recent computational method, termed Beyondcell, that identifies tumor cell subpopulations with distinct drug responses from single-cell RNA sequencing (scRNA-seq) data for proposing cancer-specific treatments, have focused on pre-clinical models but lack validation in patients at the clinical level^{24–28}. Additional efforts to identify biomarkers of response and resistance at the patient level using sc-expression are emerging for both targeted therapies and immunotherapies, with remarkable results^{29–31}. However, to date, harnessing the sc tumor transcriptomics of patients for tailoring their treatment in a direct, systematic manner has remained an important open challenge.

Aiming to address this challenge, here we present a precision oncology framework for PERsonalized Single-Cell Expression-based Planning for Treatments In ONcology (PERCEPTION). This approach builds upon the recent availability of large-scale pharmacological screens and sc-expression data in cancer cell lines to build machine-learning-based predictors of drug response based on the gene expression of single cells. We first show that PERCEPTION can predict the response to single and combination treatments in three independent screens performed in cancer and patient-tumor-derived primary cells, based on their sc-expression profiles. Secondly, we show that PERCEPTION can stratify the responders versus non-responders in two cohorts, multiple myeloma and breast cancer, with patients' tumor sc-expression profiles and can capture the development of resistance using longitudinal tumor sc-expression profiles during treatment in a cohort of lung cancer patients. Notably, PERCEPTION markedly outperforms state-of-the-art sc-based and bulk-based predictors in all three sc clinical cohorts considered. Finally, we provide a guide for using PERCEPTION for a new clinical cohort with sc-expression to select patients for receiving treatment. In sum, we present a computational approach that showcases the exciting potential of sc-gene expression-based precision oncology.

Results

Overview of PERCEPTION

To predict patient response to therapy from the tumor's sc-expression profile, we built a three-step machine-learning pipeline called PERCEPTION (Fig. 1a; a detailed description is provided in the Methods). One of the key challenges in building a supervised machine-learning model to predict clinical response using sc-expression is the lack of large-scale sc-expression data with clinical response labels. To overcome this issue, we used the concept of transfer learning, a machine-learning technique where a model trained on one task (for which considerable data are available) is used as the starting point for a model on a second, related task for which less training data are available. Transfer learning allows the second model to benefit from the knowledge learned by the first model.

We built PERCEPTION response prediction models for each drug in three steps. In step one, a bulk-expression model is trained to

predict drug response in cell lines from the large-scale bulk-RNA-seq. In step two (tuning), the bulk-expression models are tuned using the cell-lines' sc-expression and drug response to build sc-expression models. Finally, in step three, we identify a heuristic strategy to predict clinical response by analyzing a clinical cohort with treatment response and sc-expression. For a given drug, we provide the input of its drug response and matched bulk-expression in cell lines for the first step, matched drug response and sc-expression in cell lines, and finally, sc-expression of the patient in the third step.

To gather cell line data for building the predictors in the first two steps, we mined bulk-expression³² and drug-response profiles (PRISM) of 488 cancer cell lines (Supplementary Table 1) from the DepMap database³³. The sc-expression profiles of these cell lines ($n = 205$; Supplementary Table 1) were obtained from a previous publication³⁴. Drug efficacy (also referred to as viability) is measured by the area under the curve (AUC) of the viability–dosage curve, where lower AUC values indicate increased sensitivity to treatment (Supplementary Table 1).

For a given drug, PERCEPTION uses the above data to build a drug-specific response predictor in cell lines through the following two steps. Step one involves building bulk-expression models. We first build a linear model with elastic net regularization of drug response using the bulk-expression and drug-response data available for 318 PRISM cancer cell lines from 21 cancer types (Extended Data Fig. 1a). Step two involves building sc-expression models. The goal of this step is to build sc-expression-based prediction models of drug response. To this end, we determine the number of genes used as predictive features (hyperparameter tuning) that maximize the ability to predict the response from sc-expression data, analyzing the 169 cancer cell lines for which both scRNA-seq profiles and drug response data are available (Extended Data Fig. 1b). To evaluate the performance of an sc model in a cell line, PERCEPTION predicts the response to a given drug for each of its individual cells, and the mean response over all those individual cells is taken as the predicted sc-based response of that cell line to that specific drug. The output of this machine-learning pipeline is hence a drug-specific sc-response model and a quantification of its predictive accuracy from sc-expression in cell lines. We evaluate this model's performance in an unseen test subset of the cell lines, using a standard leave-one-out (one cell line) cross-validation procedure. As described in the Methods, the models for some drugs will be deemed sufficiently predictive and the models for other drugs will not. Only drugs with predictive models from step two are considered in step three.

In the third and final step, we predict the clinical response in patients, which is the ultimate goal of our study. This is done using the following heuristic procedure: (1) We first identify the major cancer cell clusters in the patient's tumor using the sc-expression (transcriptional clones, a cluster of single cells whose transcription profile looks similar). (2) We then compute the mean expression of each transcriptional clone and use this as an input to the predictive drug-specific models yielded from step two to predict drug response for each transcriptional clone (if a combination of drugs is used in the treatment, we take the maximum predicted killing among those drugs as the predicted killing effect on that clone, following the independent drug action (IDA) principle³⁵). (3) Finally, the overall patient response is predicted as the minimum response among all clones, taking the stance that the clone predicted to be most resistant will likely determine the overall clinical response. As we describe later in the Results, this prediction strategy was determined in a cohort of patients with multiple myeloma by studying five different potential strategies and was then fixed and applied as-is to two other patient cohorts. For any new drug in a new cancer type cohort, the response model (steps 1 and 2) should also be built using all cell lines available in the screen (pan-cancer model), as we found that this pan-cancer model construction performs better during cross-validation than building cancer-type-specific models that use cell lines belonging only to the patient's cancer type (Extended Data Fig. 1c).

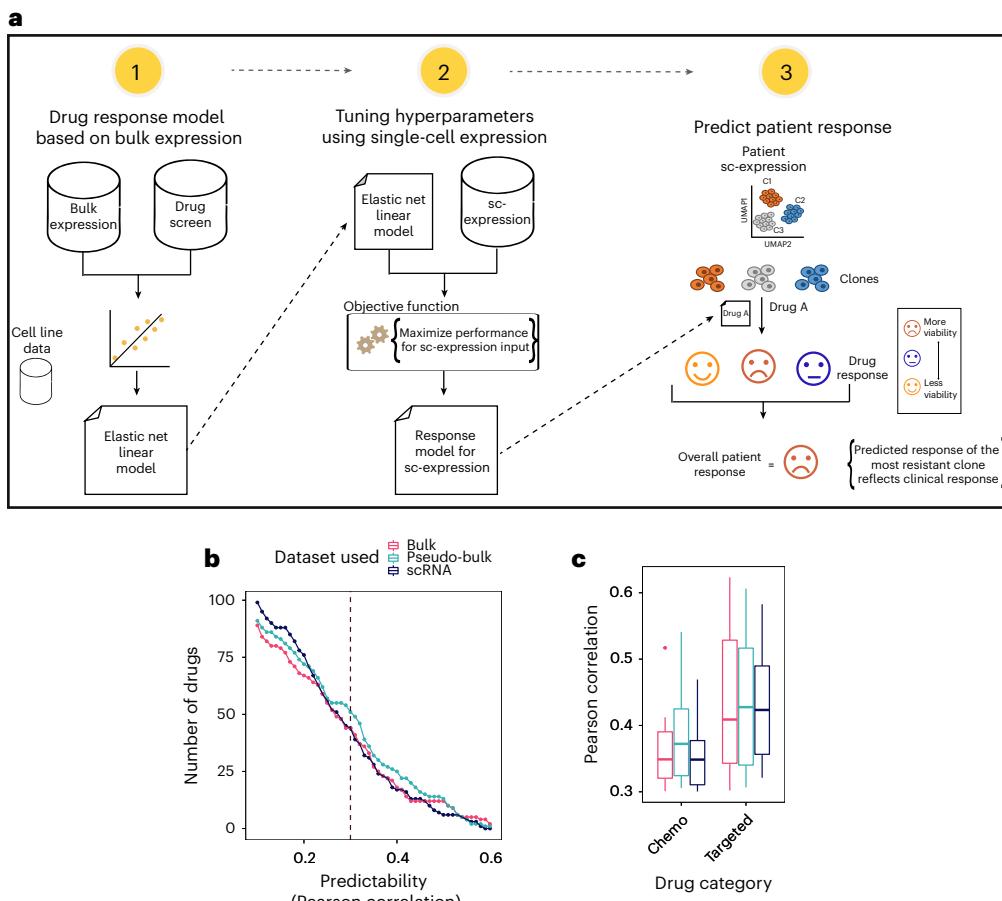


Fig. 1 | Overview of the PERCEPTION framework and its performance during cross-validation. **a**, PERCEPTION builds drug-specific models in three steps. (1) Bulk-expression response models are built based on drug response data measured in large-scale drug screens performed on cancer cell lines and their matched bulk expression. (2) Then sc-expression models are built by tuning the bulk-expression models, determining the optimal number of genes used as predictive features that maximize its prediction performance based on sc-expression of cancer cell lines. (3) In the third and final step, the clinical response in patients is predicted following a three-step heuristic procedure. Given a patient's scRNA-seq data from the tumor, identify the major cancer cell clusters (called a transcriptional clone) and their mean expression. Use this mean expression as an input to the PERCEPTION model built in step two, yielding a

predicted drug response for each transcriptional clone separately. The minimum response among all clones is predicted to be the patient's response. **b**, The number of PERCEPTION predictive models of FDA-approved drugs (y-axis), when built from sc-expression (blue), bulk-expression (red), and pseudo-bulk, as a function of the Pearson correlation between predicted and observed response values (x axis, the dashed vertical line denotes the 0.3 threshold selected). **c**, The distribution of predictive performance (x axis) of the models. In the boxplots, the center line, box edges and whiskers denote the median, interquartile range and the rest of the distribution, respectively, as in standard boxplots. Interestingly, the predictive performance is overall considerably higher for targeted therapies than for chemotherapies. A two-sided Wilcoxon rank-sum test was performed to compare groups with $n = 44$ drugs.

We test and demonstrate PERCEPTION's performance in predicting the response to monotherapy and combination treatments in screens performed in cancer and patient-tumor-derived primary cells. Then, focusing on patient data as the main goal of this investigation, we study its ability to predict treatment response in two clinical cohorts and to predict the emergence of resistance in a third clinical cohort. We additionally compare PERCEPTION's prediction performance with published state-of-the-art sc-based and bulk-based methods. Finally, we provide a guide for using PERCEPTION to predict responses in new datasets.

Cross-validation and independent performance in cell lines

We applied PERCEPTION to build response models for 133 US FDA-approved oncology drugs tested in the PRISM drug screen (Supplementary Table 2 and Extended Data Fig. 1d) and computed their performance to predict response in a leave-one-out cross-validation and tenfold cross-validation. Prediction performances for each of these drugs are provided in Fig. 1b. We deemed models to be sufficiently predictive if the Pearson correlation between their predicted

(mean sc-response per cell line) versus the observed viability on the test data was greater than 0.3. This threshold was chosen as it corresponds to the mean cross-screen replicate correlation observed among three major pharmacological screens and confirmed by us, as well as being previously reported (average cross-platform correlation across GDSC³⁶, CTD³⁷ and PRISM³⁸ is -0.30). We were able to build predictive models for 33% of the drugs tested (44 out of 133 drugs; Supplementary Table 2 and Fig. 1b). The mean performance of PERCEPTION's leave-one-out cross-validation and tenfold cross-validation are 0.39 and 0.36 for 44 drugs with predictive models (Fig. 1b). Studying this subset, in which we are able to build predictive models, we found that the drugs in this subset are more likely to be targeted therapy (mean Pearson's Rho, 0.43 vs 0.35 for chemo) and have a higher variance in response profile (Wilcoxon $P = 5 \times 10^{-7}$) and bimodality index in their response profile during training (reflecting the presence of both sensitive and resistant cell lines).

Studying the predictive accuracy of these 44 predictive models in a cross-validation manner for different kinds of transcriptomics inputs, including sc-expression, bulk-expression and pseudo-bulk-expression

(generated by summing the gene-mapped reads across single cells; Methods), we reassuringly find that the predictive performance of PERCEPTION for sc-expression as inputs on these cell lines is comparable to the performance obtained using bulk-expression or pseudo-bulk as inputs (Fig. 1c). Importantly, we note that a model built on only scRNA-seq without any pre-training on bulk-RNA-seq has markedly lower prediction accuracy (Pearson's Rho, 0.22 vs 0.39 for the 44 predictive drugs in the left-out test cell lines), highlighting the importance of pre-training on bulk. We visualized PERCEPTION's predicted killing levels at sc resolution for eight FDA-approved drugs with high-confidence mechanisms of action and the activity of each pathway they are targeting (Extended Data Fig. 2).

We next asked what are the identities of the genes that these 44 models are using to predict drug response. An average of 76 genes are used as features in the above models after regularization, in which the key pathways enriched include apical junction pathway, including genes like *ABCB1*, encoding multi-drug resistance 1 (MDR1) a transporter implicated in resistance to many drugs, cell-cycle-related targets and more (Extended Data Fig. 1e).

We next evaluated PERCEPTION's performance on three independent large-scale cell-line screens, two cultured (Nair³⁹ and GDSC) and one PDC, to stratify the resistant versus sensitive cell lines (top vs bottom 33% by viability, respectively). We built PERCEPTION models for each drug across the three screens individually. We note that we were unable to build predictive models for any drugs in the PDC screens using PRISM data and therefore used GDSC data (~800 cell lines). Detailed methods on how PERCEPTION models were built and used are provided in the Methods. PERCEPTION was able to stratify the resistant versus sensitive cell lines with an average AUC under receiving operator curve (henceforth referred to as AUC) of 0.81 (AUC = 0.87 for cultured cell lines, Extended Data Fig. 3a–g; AUC = 0.75 for PDCs, Extended Data Fig. 3h–k). A detailed performance evaluation including drug-level performance measures is provided in Extended Data Figs. 4 and 5. Predicted and observed viabilities are also strongly correlated in all three datasets (Pearson's Rho, 0.36 for Nair, 0.28 for GDSC and 0.64 for PDCs; Extended Data Figs. 5 and 6). We note that a control PERCEPTION model that is not tuned on sc-expression yielded a modestly inferior performance in this test (average AUC = 0.71, for cultured cell lines AUC = 0.81, for PDCs AUC = 0.62).

Predicting treatment response in a multiple myeloma trial

After showing that PERCEPTION's cell-line-based model can predict the response of monotherapy and combination in cultured and PDC lines, we next ask how we can use the cell-line-based models to predict patient response using the pre-treatment sc transcriptomics from their tumors. To this end, we mined the largest such dataset published to date, including data from 41 patients with multiple myeloma. The patients were treated with a DARA-KRD combination of four drugs: daratumumab (monoclonal antibody targeting CD38), carfilzomib (proteasome inhibitor), lenalidomide (immunomodulator) and dexamethasone (anti-inflammatory corticosteroid)²⁹. The sc-expression and clonal (transcriptional cluster) composition and treatment response labels, as determined in the original study²⁹, were available for 28 tumor samples from these patients (Fig. 2a). Patient response was measured by tumor size estimates in radiological images.

As explained above in the PERCEPTION overview, to predict the clinical response from a tumor's sc-expression, PERCEPTION first finds the major transcriptional clones (provided in the original publication²⁹) and predicts the treatment response for each clone separately (response is defined as the predicted reduction in viability after treatment; see Methods). Figure 2b shows the predicted viability of the combination at a clonal level for each patient. We designed and tested five different strategies to predict the clinical response from clone-level killing to find the most optimal strategy. We tested their performance for stratifying responders ($n = 7$) versus

non-responders ($n = 21$) (Fig. 2c, see Methods). In brief, the clinical response of a patient is determined by computing one of the following strategies: (1) weighted average response, an average of response across all the clones weighted by their abundance in the tumor; (2) unweighted average response, an average of response across all the clones; (3) most-sensitive clone response, the response of the most-sensitive clone; that is, the clone with the highest predicted response; (4) unweighted most-resistant clone response, the response of the most-resistant clone; that is, the clone with the least response; (5) most-resistant clone response, the response of the most-resistant clone, weighted by its abundance proportion. The resulting AUCs for these strategies were 0.59, 0.55, 0.64, 0.75 and 0.83, respectively (Fig. 2c). This analysis revealed that the fifth strategy best predicts the clinical response. In cell lines, this strategy also stratified resistant versus sensitive, albeit with lower performance (AUC = 0.79; Extended Data Fig. 6i) than the mean-response strategy (AUC = 0.89).

As an illustrative example using the most-resistant clone strategy (Fig. 2b), in a sample from patient Kydar19, there are three clones: c1, c2 and c3. Here, c2 and c3, two low-abundance clones, are predicted to be relatively responsive to the treatment, whereas c1, the most abundant clone, is predicted to be resistant. In this case, c1 is likely to drive the patient response, and thus, the patient will be predicted to be resistant or have a low response to the treatment. The resulting predicted response scores from this strategy are significantly higher in responders versus non-responders (Fig. 2d), successfully predicting the treatment response (AUC = 0.83; Fig. 2e). This may be the case because the most resistant clone is most likely to be selected upon treatment and end up dominating the tumor, thus best reflecting the clinical response. From here onwards, we fixed this most-resistant clone response strategy for predicting clinical response and tested it in two additional cohorts. The top pathways enriched among the gene features used by the PERCEPTION model are surfactant metabolism and O-linked glycosylation of mucins.

Predicting CDK inhibition response in a breast cancer trial

Using the most-resistant clone response prediction approach described in the previous subsection, we next tested PERCEPTION's ability to predict patient response in the FELINE breast cancer clinical trial⁴⁰. This clinical trial includes three treatment arms: endocrine therapy with letrozole (arm A), an intermittent high-dose combination of letrozole and CDK inhibitor ribociclib (arm B) and a continuous lower dose combination of the latter (arm C). sc-expression and treatment response labels were available for 33 patients (arms A, B and C having 11 samples each; Supplementary Table 7). Patient response was determined by tumor growth measurements from mammogram, MRI and ultrasound of the breast.

We could build a (borderline) predictive PERCEPTION response model for only the CDK4/6 inhibitor ribociclib (Pearson's $R = 0.26$, $P = 1.5 \times 10^{-3}$), and therefore we focused our analysis on the combination arms B and C that include it (Fig. 3a). We processed the sc-expression profiles of the tumor cells as previously described⁴⁰ and identified 38 transcriptional clusters or clones that are shared across the patients (Extended Data Fig. 7a–c; see Methods). Patient response was predicted based on the pre-treatment samples, following the same strategy used in the multiple myeloma case. As the number of patients in each arm (B and C) is quite small, we predicted the response of the patient pre-treatment samples in aggregate. The resulting predicted viability of the non-responders is higher than that of the responders (Wilcoxon rank-sum test, one-sided $P = 0.05$; Fig. 3b), as expected. PERCEPTION successfully stratified the responders versus non-responders with an AUC of 0.776 (Fig. 3c). Aligning to our known mechanism of action of ribociclib's inhibition of CDK4/6 activity, leading to cell cycle arrest, PERCEPTION's signature comprising 72 genes is enriched in pathways involved in cell cycle, specifically, TNF receptor family ($P = 0.004$) and regulation of p53 ($P = 0.004$).

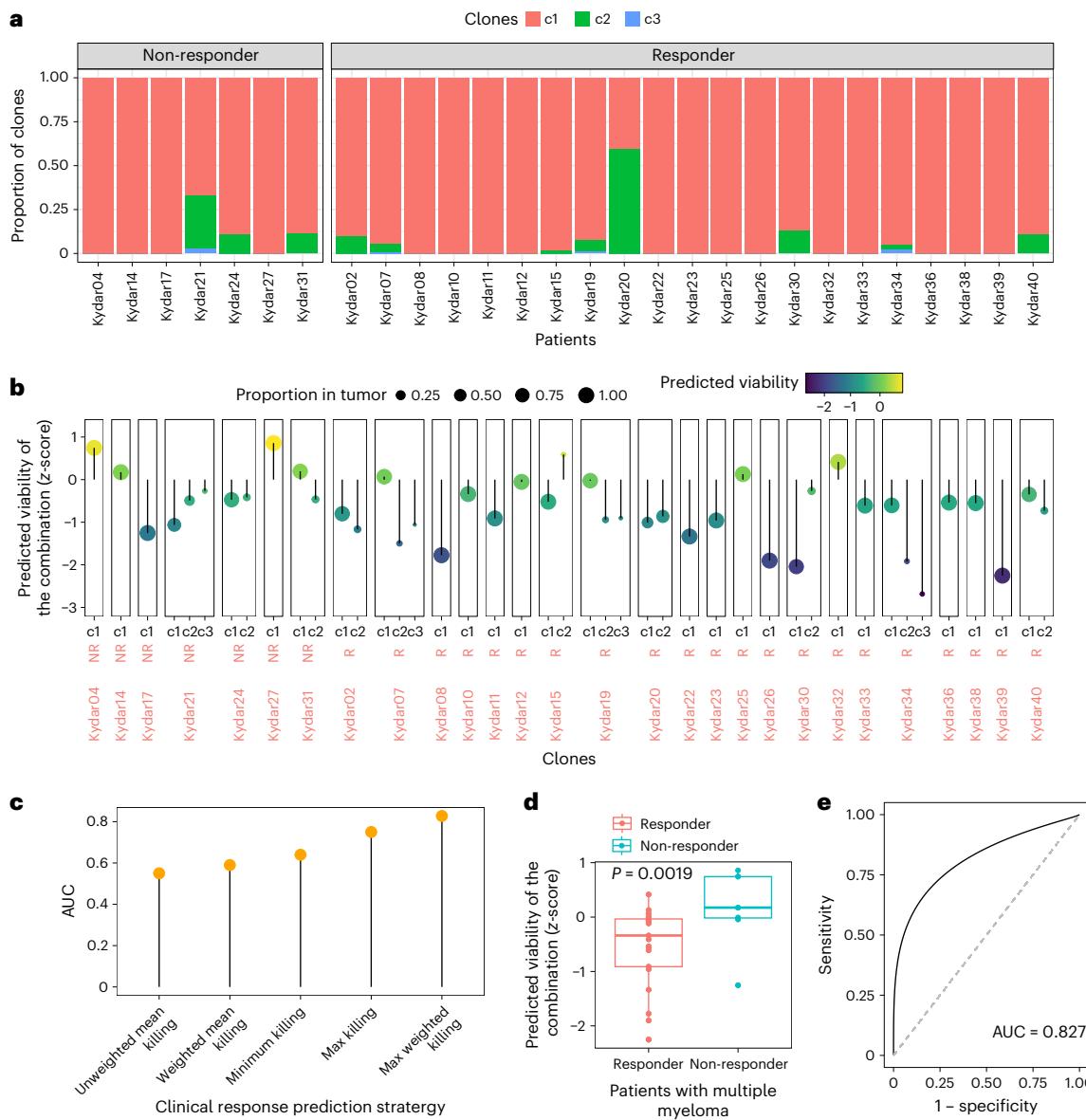


Fig. 2 | PERCEPTION predictions of DACA-KRD combination therapy in patients with multiple myeloma. **a**, Distribution of abundance of the transcriptional clones (yaxis) in each patient with multiple myeloma (xaxis); the color code for the clones is provided at the top. **b**, Predicted viability of the combination at a clonal level for each patient; the response status is provided at the bottom strip of each facet. The left-to-right order of patients is the same as in panel **a**. **c**, The stratification performance in distinguishing responders versus non-responders from the clone-level predicted response information (yaxis) of five different strategies (xaxis). **d**, The predicted combination response in

28 patients with multiple myeloma stratified by responder ($n = 21$) versus non-responder ($n = 7$) status. A two-sided Wilcoxon rank-sum test was performed to compare groups. The boxplot shows median (center), 25th and 75th percentiles (bounds of box) and minima and maxima (whiskers). **e**, Receiver operating characteristic curve displaying the predicted combination response; AUC denotes the overall stratification power in distinguishing responders versus non-responders. The gray dashed line represents the line of no discrimination, which illustrates the performance of a purely random classifier with an AUC of 0.5.

Capturing emergence of resistance in lung cancer patients

We next tested whether PERCEPTION can capture the development of clinical resistance during targeted therapy treatment in patients. To this end, we analyzed a published cohort with scRNA-seq profiles of 24 patients with non-small cell lung cancer (NSCLC) with 14 pre-treated and 25 post-treated biopsies (Extended Data Fig. 8a–f and Supplementary Table 8)⁴¹. In total, patients in this cohort were treated with four different tyrosine kinase inhibitors, including erlotinib (a first-generation EGFR inhibitor), dabrafenib (a serine/threonine kinase inhibitor), osimertinib (a third-generation EGFR inhibitor) and trametinib (a MEK inhibitor). Based on the notion that the resistance to these targeted therapies frequently increases as the treatment time grows, we reasoned that the predicted response for a

given post-treatment biopsy would decrease (reflecting an increase in resistance to that treatment) as time elapses from the treatment start.

To test this hypothesis, for each post-treatment biopsy, we defined its estimated ‘extent of resistance’ to a given treatment as the difference between its PERCEPTION-predicted response and the baseline predicted response. The latter was computed as the mean predicted viability across all pre-treatment biopsies (as the majority of the samples were not matched, precluding an overall pairwise matched comparison). We found that the extent of resistance to treatment increases with the elapsed time since the start of treatment, but only in those patients reported to acquire resistance (progressive disease; Spearman’s Rho = 0.634, $P = 0.026$, $n = 17$; Fig. 4a). We also found that this positive correlation between the elapsed treatment time and the estimated

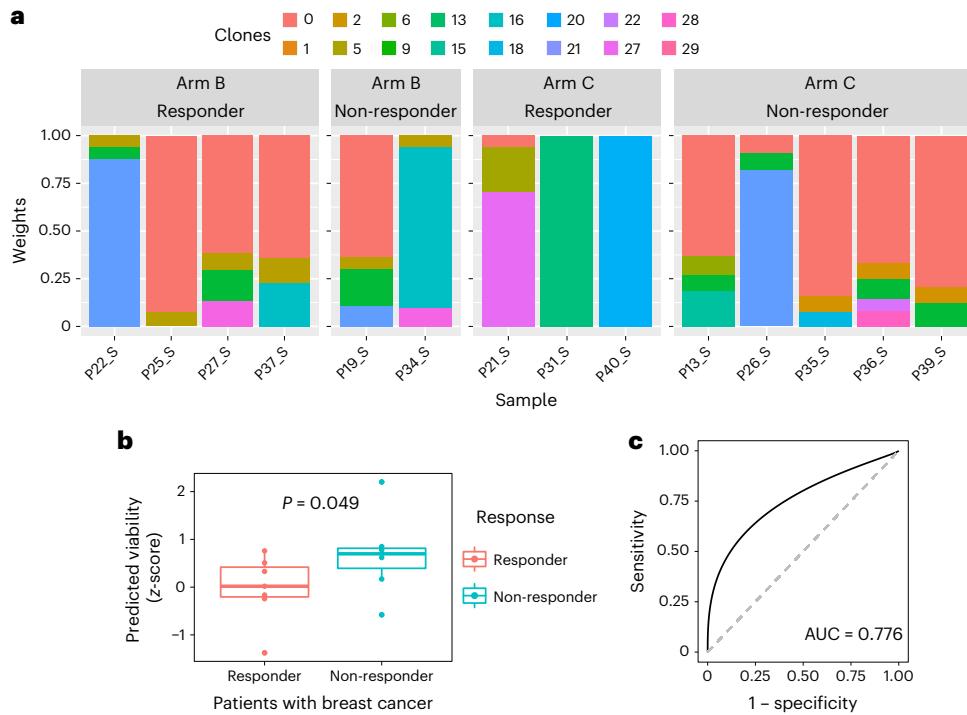


Fig. 3 | PERCEPTION prediction of the combination therapy in the FELINE clinical trial. **a**, Transcriptional clone composition (yaxis) in each tumor from a patient with breast cancer studied in the combination arms B and C (xaxis); color code for the clones is provided at the top. On the xaxis, the labels are a combination of the patient identification and the time point at which the sample was collected (“_S”, day 0; “_E”, day 180). **b**, The predicted combination response in 14 patients with breast cancer (samples collected at day 0), stratified by their responder ($n=7$) versus non-responder ($n=7$) status. Two-sided

Wilcoxon rank-sum tests were performed to compare groups. The boxplot shows median (center), 25th and 75th percentiles (bounds of box), and minima and maxima (whiskers). **c**, Receiver operator characteristic curve displaying the predicted combination response; AUC denotes the overall stratification power in distinguishing responders versus non-responders. The gray dashed line represents the line of no discrimination, which illustrates the performance of a purely random classifier with an AUC of 0.5.

extent of resistance holds true when patients receiving different drugs are analyzed separately (Extended Data Fig. 9a), when controlling for prior treatments (Extended Data Fig. 9b), when individual patients are analyzed separately (Extended Data Fig. 9c) and when controlling for tumor stage (Extended Data Fig. 9d). The extent of predicted resistance is significantly higher in post-treatment biopsies collected from the patients with progressive disease versus residual disease (Wilcoxon rank-sum $P < 0.002$, stratification AUC = 0.88; Fig. 4b). Notably, we do not observe this strong positive correlation but rather a negative trend in patients who responded well to the treatment (residual disease, $n=7$, Spearman’s Rho = -0.67 , $P = 0.11$; Fig. 4a). The observed increase in the predicted extent of resistance to treatment with elapsed treatment time occurred specifically in patients who had acquired resistance.

We next analyzed the subset of patients with matched biopsies, including five patients with two biopsies each and one patient with four biopsies. Analyzing these samples in a matched manner, we find that the correlation between treatment elapsed time and the estimated extent of resistance holds true in the matched cases, and only in the patients who have acquired resistance (regression interaction $P = 0.003$). Of particular interest is a case of a single patient (TH179), treated with dabrafenib, who had four biopsies at two different time points and developed progressive disease. The predicted viabilities to dabrafenib of the four tumor biopsies taken after 331 and 463 days of start of treatment are significantly higher than pre-treatment biopsies (Fig. 4c). Furthermore, the predicted viabilities of all three biopsies from day 463 are significantly higher than the biopsy from day 331. Notably, we find that the abundance of the top 50% of predicted resistant clones increases while the abundance of the bottom 50% of predicted resistant clones decreases with the elapsed time since the start of treatment, as one would expect (Fig. 4d and Methods).

The rate of increase of abundance is significantly higher in the top 50% of the predicted resistant clones than in the bottom 50% (Fig. 4d and Methods). Taken together, these results testify that PERCEPTION can capture and quantify the emergence of treatment resistance as the disease progresses.

We next found that the features or genes used by the above models are enriched in pathways involved in cell junction organization and cell–cell communication, including extracellular matrix organization, RHO GTPase cycle, and NOTCH signaling. We also found that this signature is enriched in the recently reported resistance mechanism for EGFR-inhibitors (EGFRi) via hypermutators driven by AXL⁴². Specifically, our prediction signature is enriched in the three resistance pathways identified in that study⁴²: AXL overexpression signature ($P = 3 \times 10^{-3}$), MYC overexpression ($P = 2.1 \times 10^{-2}$) signature and purine synthesis ($P = 1.6 \times 10^{-4}$).

To prioritize candidate drugs available in this cohort whose treatment may overcome the resistance acquired, we asked whether the development of resistance to a drug can induce either cross-sensitivity or cross-resistance to the other drugs⁴³. We focused on the patients (Supplementary Table 8) who acquired resistance and computed the PERCEPTION response predictions for each of these drugs and the correlations between these drug sensitivity predictions across these patients (Fig. 4e and Methods). PERCEPTION predictions suggest that the development of resistance to erlotinib would induce a cross-sensitivity to gemcitabine (Fig. 4f, top-left panel; Pearson’s $R = -0.94$, $P = 0.06$) and cross-resistance to dabrafenib (Fig. 4f, top-left panel; Pearson’s $R = 0.91$, $P = 0.09$). A literature survey (Methods) revealed that gemcitabine treatment can overcome erlotinib resistance in cancer cell lines through the downregulation of Akt⁴⁴. In patients, a combination of gemcitabine + erlotinib in pancreatic

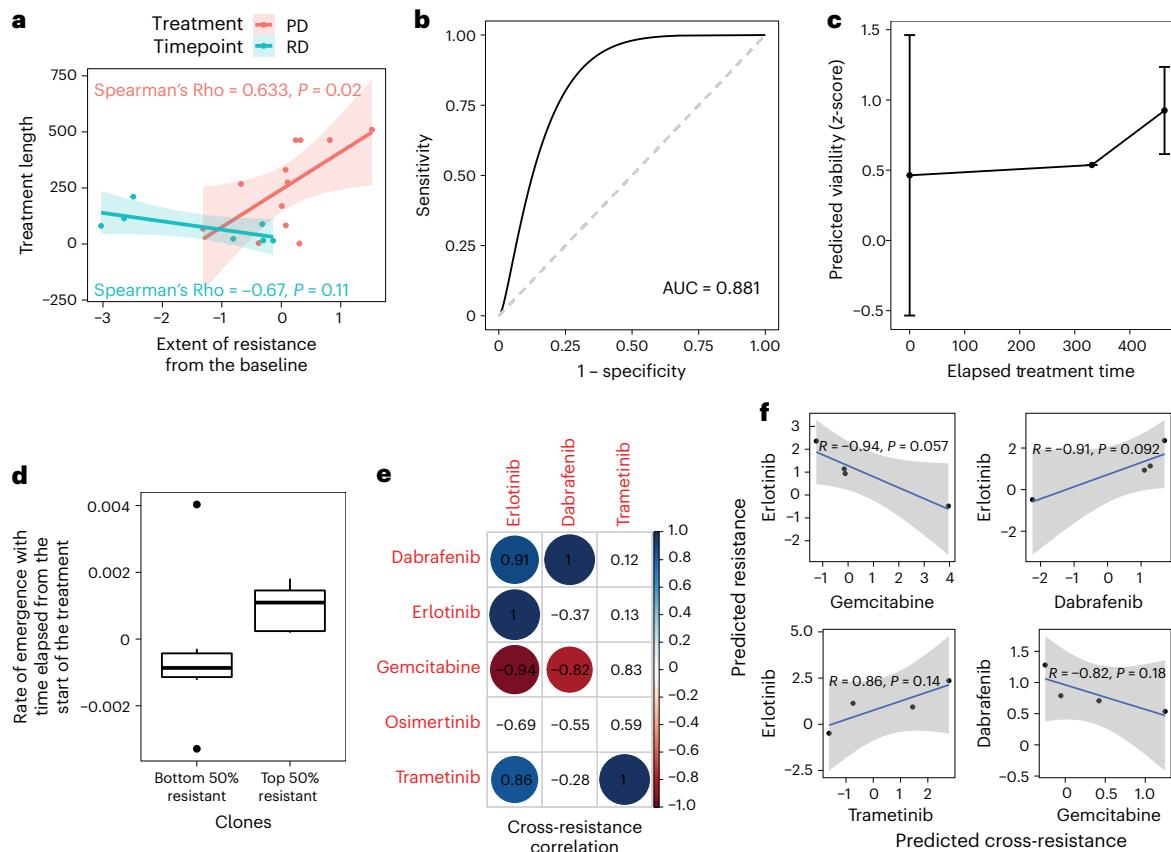


Fig. 4 | Predicting the development of resistance to tyrosine kinase inhibitors in lung cancer patients. **a**, The extent of predicted reduced killing (as a corollary of resistance) to a treatment from the baseline (*x* axis) is correlated with the time elapsed (days from start of treatment until biopsy) (*y* axis). The points and line colors denote whether the biopsy is from patients with progressive disease or from responders. Error bars, 95% confidence intervals of the fit. **b**, Receiver operating characteristic (ROC) curve depicting PERCEPTION predictive power in distinguishing progressive ($n = 17$) versus responding ($n = 7$) patients. The gray dashed line represents the line of no discrimination, which illustrates the performance of a purely random classifier with an AUC of 0.5. **c**, The case of patient TH179 with multiple biopsies is presented, where the predicted viability in 14 pre-resistant (day 0) and 4 post-resistant tumors at day 331 ($n = 1$) and day 463 ($n = 3$) to dabrafenib is shown. Error bars, minimum and maximum values owing to the small sample size of three data points. **d**, The rate of change in abundance of top versus bottom 50% predicted resistant clones ($n = 21$ each)

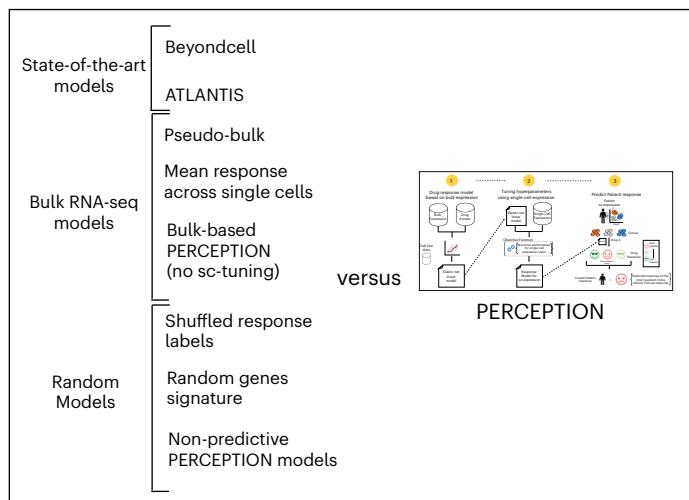
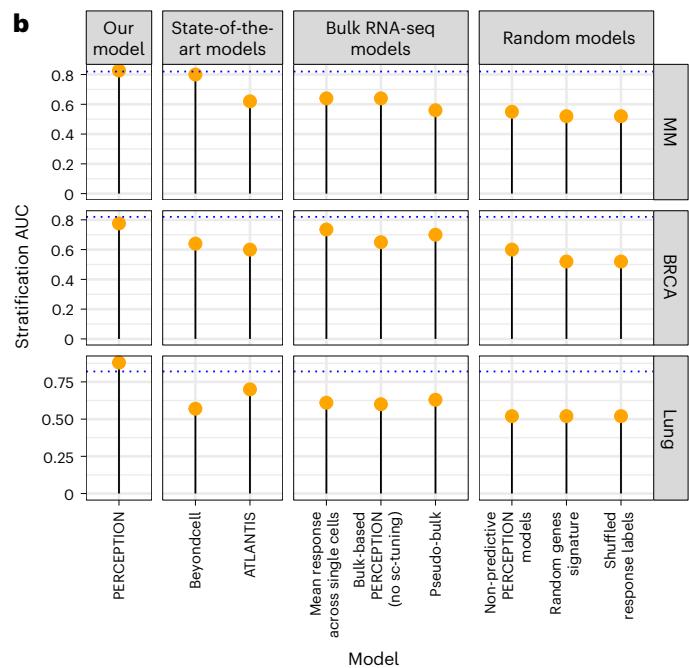
with elapsed time since the start of treatment. Boxplot shows median (center), 25th and 75th percentiles (bounds of box), and minima and maxima (whiskers). **e**, Correlation matrix of the extent of resistance among drugs available in the trial across all patients who have acquired resistance. The strength of the correlation (Pearson's R) is provided in the respective box, represented by the size of the circle, where the color represents whether the correlation coefficient is negative or positive (red and blue, respectively). This is computed for drugs with at least three resistant patients (no. of patients, 4, 4 and 3, respectively). The drugs with correlations of $P < 0.1$ (before false discovery rate correction) are indicated by colored circles. In both **c** and **d**, two-sided Wilcoxon rank-sum tests were performed to compare groups. **f**, Correlation matrix illustrating cross-resistance between various drugs. The matrix represents the results of our analysis to identify pairs of drugs A and B, where resistance to drug A may induce cross-resistance to drug B. The cross-resistance relationship can be asymmetric. The P value associated with Pearson's R denotes correlation significance.

cancer in phase III trial has shown higher overall and progression-free survival^{45,46}. By contrast, the addition of trametinib to erlotinib did not significantly improve survival in a phase I/II clinical trial⁴⁷. In sum, our analysis supports the possibility that erlotinib resistance may induce cross-sensitivity to gemcitabine, which may be of interest for future testing.

Predicting combination therapies targeting disjoint clones

We next turned to investigate PERCEPTION's capability to identify effective combination treatments in clinics. To this end, we curated clinical trial data of various combinations tested for NSCLC with survival information to assess the predictive power of PERCEPTION models. The trial data were curated from TrialTrove (Methods). We found that PERCEPTION's predicted improvement in response to combinations versus the pertaining monotherapies is correlated with the survival improvement owing to the combination observed in the respective clinical trials (see Extended Data Fig. 10a–c for multiple myeloma and Extended Data Fig. 10e–k for lung cancer; weighted Pearson's $\text{Rho} = 0.66, P = 0.02$, weighted by the number of patients in a trial).

The only targeted therapy with enough unique combination trials is erlotinib, and repeating this analysis for erlotinib yielded concordant results (Pearson's $\text{Rho} = 0.76, P = 0.08$; Extended Data Fig. 10h). Aside from the trials tested, among all possible combinations tested of approved drugs, the top-ranking pathways composing combinations pairs are the tyrosine kinase pathway and the tubulin polymerization pathway (Extended Data Fig. 10i–k). This analysis was also done for multiple myeloma, and the results are presented in Extended Data Fig. 10a–d. Our top-ranked combination pair is niraparib and ponatinib, an EGFR inhibitor and a canonical BCR-ABL inhibitor, respectively (disjoint killing score (DKS) = 0.25, empirical $P = 1 \times 10^{-4}$). The next top combination pair with the high DKS is lapatinib and thioguanine (DKS = 0.24, empirical $P = 1 \times 10^{-4}$), a dual HER2 and EGFR inhibitor and a purine inhibitor, respectively. Analogously, we next looked for all possible triplets of drug combinations exhaustively (Extended Data Fig. 10b; $n = 13,244$). Our top hits include the combination of gefitinib, icotinib and trametinib (DKS = 0.21, empirical $P = 1 \times 10^{-4}$) and gefitinib, lapatinib and trametinib (DKS = 0.19, empirical $P = 1 \times 10^{-4}$) (Extended Data Fig. 10d).

a**b****Fig. 5 | Performance of PERCEPTION versus state-of-the-art and RNA-seq models.**

a, Illustration of our overall comparison comparing the PERCEPTION model (on the right) versus eight other models, including two state-of-the-art sc models (Beyondcell and ATLANTIS), three bulk-RNA-seq-based models (pseudo-bulk, mean response across all single cells and PERCEPTION trained on only bulk-model without sc-training) and finally three random models. **b**, The stratification

performances of differentiating responders and non-responders are provided on the y-axis for each model versus the PERCEPTION in three clinical cohorts: MM (multiple myeloma, 21 responders, 7 non-responder), BRCA (breast cancer, $n = 7$ each of responders and non-responders) and lung cancer (17 progressive and 7 responding). The dotted blue line denotes the mean performance of PERCEPTION across the three cohorts.

Benchmarking PERCEPTION versus state-of-the-art methods

We compared the prediction performance of PERCEPTION in the above three clinical cohorts versus two different published predictors and four other alternatives that we implemented (Fig. 5a): (1) a state-of-the-art model based on sc-expression (Beyondcell²⁷); (2) a state-of-the-art bulk-expression-based model (ATLANTIS³³); (3) usage of pseudo-bulk-RNA-seq (Pseudo-Bulk); (4) taking the mean viability across all single cells in a tumor sample (mean viability, the strategy we used for predicting response in cell lines and PDCs (mean-response-sc)); (5) Bulk-based-only PERCEPTION models that are not tuned on sc-expression; and (6) three kinds of random models created using shuffled viability labels, random gene signatures and random coefficients (Methods). Notably, across the three cohorts as well as in each individual cohort, PERCEPTION was the best-performing model by a considerable margin (mean AUC = 0.828; Fig. 5b) compared to the published state-of-the-art methods. The other models studied here achieved mean AUCs as follows: state-of-the art models (Beyondcell, 0.67; ATLANTIS, 0.64); the three bulk expression-based models that we generated (Pseudo-bulk, 0.63; mean viability, 0.663; bulk-based-only PERCEPTION models, 0.63) and finally, three randomly generated models (as expected, shuffled viability labels, 0.51; random gene signature, 0.55; random coefficient model, 0.53). Notably, across the three clinical cohorts studied, the mean AUC improvement of PERCEPTION over the previous best-published model, Beyondcell, is considerable (0.15, $P = 0.002$).

How to use PERCEPTION for new a cohort or a new drug

We provide predictive pan-cancer drug models for 44 FDA-approved drugs in our source data. For a new clinical trial dataset with sc-expression that involves drugs with existing predictive models, PERCEPTION can be run using a single script (Running_PERCEPTION_for_new_dataset) in our GitHub repository.

When the drugs involved do not have given predictive models, one can still aim to build PERCEPTION models, as follows. First, this

process requires the following two inputs: sc-expression of cancer cells from the tumor and treatment information. Second, the process involves three steps: step one: the user should first build a bulk-expression-based model for the given treatment. One can readily aim to build models for any of the 1,500 drugs currently available in DepMap. We recommend that the user only consider using models that surpass the predictive threshold we used (Pearson's correlation of >0.3 between observed and predicted). We also recommend training such models on all cell lines available (pan-cancer model) versus training on the subset of cell lines from the pertaining specific cancer type of the patient's cohort, as we found that pan-cancer models perform better in both patients and cell lines (Extended Data Fig. 1c; AUC = 0.75 vs 0.88 and AUC = 0.52 vs 0.77 in lung and breast cohort, respectively; decrease of Pearson's correlation of 0.38 to 0.25 in cell lines). A similar approach and guidelines should be applied for building sc-based models. Step two: the user will next cluster the cancer cells available from the tumor, identify each cluster mean expression in the default setting and rank-normalize it. Step three: based on the sc-models, the user can now predict patient response using the three-step heuristic approach described in previous sections. The resulting response scores are predicted to stratify patients who are more likely to respond to the given treatment, whereby the higher the score, the higher the likelihood of response. The code for building and testing models for new drugs is provided (Running_PERCEPTION_for_new_dataset (mode 2)).

Discussion

We present PERCEPTION, a computational pipeline for systematically predicting patient response to cancer drugs at sc resolution. We demonstrate its application for predicting response to monotherapy and combination treatment at the level of cell lines and PDCs as well as in predicting patient response in three recent sc clinical cohorts spanning multiple myeloma, breast cancer and lung cancer. We find that incorporating the transcriptional clonal information of the tumor

into the prediction process improves the overall accuracy. For a given patient, the transcriptional clone with the worst response (that is, the most resistant pre-treatment clone) best explains the overall response to treatment. Performing an extensive and systematic comparison with other expression-based models, we show that PERCEPTION achieves markedly superior performance compared to two previously published methods.

The observation that the most-resistant-clone strategy (the one used for predicting the response in clinical trials) can also stratify resistant versus sensitive in cell lines, albeit with lower power than the mean-response strategy might be because the clinical responses are measured at much longer time scales with the patients (months) than in the cell lines (within days). Passage of more time is probably better for the selection of the most-resistant clone. This underscores the importance of considering the repertoire of a given tumor's transcriptional clones in predicting its response to therapy. Furthermore, the observation that pseudo-bulk and bulk-based models performed better than scRNA-seq-based models in cell lines during cross-validation might be because of the relative homogeneity of cell lines, whereby scRNA-seq may not offer advantages over bulk-seq, sometimes resulting in comparable or worse predictions.

Our study's limitations include the use of homogeneous 2D cell lines and sparse pre-treatment sc datasets with response labels to train our models. As data availability increases, so will our predictors' accuracy and scope. Hence, the current demonstration of their potential value will hopefully serve to drive the generation of more pre-treatment sc datasets with clinical annotations in the future. Given the US\$150,000 average yearly cost for cancer treatment in the US⁴⁸, US\$15,000 for tumor sequencing seems justified, despite additional costs. This option should be explored further, through more sc dataset collection and predictive model development. Another limitation of our study is that our model was learned over the *in vitro* dosages whose translation to clinical response is non-trivial, and therefore we chose the AUC measure, a response measure over multiple dosages ($n = 8$), as it is more likely to lead to a more robust approach.

PERCEPTION's predictions may be further improved by considering cancer type-specific cell lines, whenever a large number of such models become available for each cancer type. The quality of our response models depends on the quality of the sc-expression profiles available; for example, their depth, drop-out rates, and so forth. We deliberately did not impute the sc data given the recent reports that dropouts are limited to non-unique molecular identifier-based sc-expression methods and otherwise probably reflect true biological variation^{49,50}. A key limitation of our pipeline is a lack of ability to predict drug effects on immune and normal cells in the tumor microenvironment, which is needed to estimate the toxicity and side effects of different combinations. A major push to future sc-based precision oncology development will come from large-scale drug screens of drugs in non-cancerous cell lines, which are currently very scarcely available. Those cell lines will enable the construction of predictors of drug killing of non-tumor cells, using an analogous pipeline to the one presented here for tumor cells.

Finally, our results demonstrate that tracking the drug response expression in post-treatment biopsies could help follow the evolution of drug resistance at sc resolution and help guide the design of future personalized combination treatments that could significantly diminish the likelihood of resistance emergence. Going beyond patient stratification, we identify new combination therapies that different individual clonal clusters for multiple myeloma and lung cancer. However, we must note that these predictions require further validation.

In summary, this study demonstrates that the high resolution of information from scRNA-seq could indeed be harnessed to predict the treatment response of individual patients with cancer in a systematic, data-driven manner. It is our hope that the results shown will herald many more such studies, sooner rather than later. Retrospective studies

on additional clinical datasets need to be done to better assess the utility of sc prediction approaches like PERCEPTION and its accuracy before it may be studied prospectively.

Methods

Data collection

We first collected the bulk-expression and drug response profiles generated in cancer cell lines curated in the DepMap³³ consortium from the Broad Institute (version 20Q1; <https://depmap.org/portal/download/>). The drug response is measured by the AUC across eight dosages and measures, using a sequencing technique called PRISM³⁸. In total, we mined 488 cancer cell lines with both bulk-transcriptomics and drug response profiles. We next mined sc-expression of 205 cancer cell lines (280 cells per cell line) generated in a previous study³⁴ distributed by the Broad Single-cell Portal. The metadata, identification and clustering information were also mined from the same portal (https://singlecell.broadinstitute.org/single_cell/study/SCP542/pan-cancer-cell-line-heterogeneity#study-download). Data collection and analysis were not performed blind to the conditions of the experiments. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

For the multiple myeloma dataset and the breast cancer dataset, the data from all human subjects are coded from two published papers^{29,40}. For the lung cancer data, we used only previously published data⁴¹ (Supplementary Table 1). The published lung cancer data we used were obtained with informed consent from all study participants based on human subject protocols (CC13-6512 and CC17618; principal investigator, C.M.B.) approved by an IRB at the University of California, San Francisco, and based on clinical trial NCT03433469. The details of the three clinical cohorts, including trial status and endpoint extraction process, are provided in Supplementary Table 9.

The PERCEPTION pipeline

A response model for a drug is built into the PERCEPTION pipeline through two steps: learn from bulk, and optimize using sc-expression. In step three, we use the models from step two to predict response in patients.

We first divided all the cancer cell lines into two sets. The first set comprised cell lines for which bulk-expression is available and sc-expression is not available ($n = 318$), and the second set comprised cell lines for which sc-expression is available ($n = 170$). The first set is used during learning from bulk (step one, expanded below) and the second is used in optimizing using sc-expression (step two).

Step one, learn from bulk. As a feature selection step, we first identified genes whose bulk expression is correlated with a drug viability profile (using the Pearson correlation). We considered the Pearson correlation $Pc(d, g)$ between drug d and gene g as a measure of information in a gene expression profile and ranked each gene based on the strength of the correlation. Considering the top X genes, where X is a hyperparameter optimized in the next step, we built a linear regression model regularized using elastic net to predict the response to d in fivefold cross-validation, as implemented in R's glmnet⁵¹.

Step two, optimize using sc-expression. We built the above model using a Bayesian-like grid search of various possible values for X (range, 10–500), whereby the model with the best performance using an sc-expression input of 169 cell lines (left one out for testing) was chosen. Finally, we measured the model performance by leave-one-out cross-validation using the left-out cell line, which was not used in either model building or hyperparameter optimization. Here, the model is trained on all data except for one sample, which is held out for testing; that is, its viability is predicted by the model. The cross-validation process is then repeated n times, with a different sample being held out each time, on which the prediction is made. After running this n times, the Pearson correlation coefficient is calculated between the predicted and the observed drug response values for all n held-out samples.

Performance was measured using Pearson's correlation between the predicted response and the actual response.

Step three. In the third and final step of PERCEPTION, we predict clinical response in patients using a cell line-based model and sc-expression profiles of the patient's tumor. We identify the major cancer cell clusters using sc-expression, compute the mean expression for each clone and use this as input for the model to predict drug response for each clone. The overall patient response is predicted as the minimum response among all clones, as we reason that the most resistant clone will determine the clinical response. Our prediction strategy was determined through trial and error in a cohort of patients with multiple myeloma and was fixed and applied to all other patient cohorts in the study. For a given treatment, we interpret this to mean that the predicted response of the most resistant clone in the patient's tumor determines the clinical response. We converged on this strategy by using a trial-and-error approach, testing five different strategies to predict a patient's response from individual clone-level responses. This strategy is then fixed. During the comparison of PERCEPTION performance versus state-of-the-art methods, we used the following three types of random models: shuffling the viability labels in the cell lines, by randomly selected gene signatures and finally using non-predictive models of other drugs.

Description of the method and optimization formula. We used an iterative approach using elastic net regression to identify the optimal number of genes that maximize the predictive performance of our model. By performing elastic net regression with different subsets of genes, we were able to determine the optimal combination of L1 and L2 penalty hyperparameters and gene features that contribute to the best predictive performance. The objective function for each iteration remains the same as the elastic net regression:

$$\min_{\beta} \left(\frac{1}{2N} \|Y - X\beta\|_2^2 + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right) \right)$$

The process involves the following steps:

1. Select a subset of genes and form the design matrix X with that subset.
2. Perform elastic net regression using the objective function above, optimizing the hyperparameters λ and α .
3. Evaluate the performance of the model using cross-validation.
4. Repeat steps 1–3 for different numbers of genes.
5. Choose the model with the number of genes that gives the maximum predictive performance.

The final chosen model would thus have the coefficients or weights of the selected genes as parameters and would be associated with optimal hyperparameters λ and α as well as the optimal number of genes.

Data choices in step one and step two of PERCEPTION. The first step of PERCEPTION (model building) used bulk-RNA-seq of 318 cell lines to build an initial set of bulk-based models based on a large set of genes as features. The second step used scRNA-seq of 169 different cell lines to further select an optimal set of predictive features, resulting in a final set of drug-specific sc-based models. This approach was designed to make sure information would not leak between two steps, leading to overfitted performance, by building the models on two entirely disjoint sets of cell lines. For some cell lines, both scRNA-seq and bulk-RNA-seq are available, and in these cases, only their scRNA-seq was used during the second step.

Evaluating PERCEPTION on three independent cell-line screens
PERCEPTION's performance on GDSC. The pharmacological drug screens performed by the PRISM and GDSC studies are based on two

independent platforms. The GDSC data were downloaded from the DepMap portal on 15 April 2020 (<https://depmap.org/portal/download>). By testing the performance of PERCEPTION on these independent screening platforms, we can measure the extent to which the expression signature captured by our drug response models can be translated across the platforms. The following steps were performed to achieve this goal:

Step one, quality check to select cell lines and drugs. Out of the 347 cell lines in common with drug response in both GDSC and PRISM, there are 120 cell lines with sc-expression data in a previous study³⁴. We considered only the drugs shared by GDSC and PRISM that have a concordant response (Pearson's Rho > 0.3, $P < 0.05$), resulting in 28 drugs. Among these 28 drugs shared between GDSC and PRISM, we were able to build predictive models for 16 drugs.

Step two, model building and parameter optimization. For each of the drugs selected above, we ran the PERCEPTION pipeline, optimizing parameters based on the sc-expression of 90 out of 170 cell lines, using the other 80 cell lines as test data.

Step three, prediction and normalization. For the drugs for which PERCEPTION could build models, we applied the models on the cell lines and obtained predictions for each individual cell. Monotherapy response for a given drug in a cell line was represented by the mean response of all the single cells ($n = 318$). Given that the range of PERCEPTION-predicted values is typically smaller than those observed in the screens (Extended Data Fig. 3g), we used scaled, predicted AUC scores (z-scores) in further analyses.

Step four, testing and performance analysis. The resulting response models were applied to the testing dataset, and the predicted AUC values were compared to the experimental responses from GDSC and PRISM. We computed two performance measures: AUC, stratifying top versus bottom 33% as resistance versus sensitive, and a correlation between predicted versus observed response. The former is provided in the main text and the latter in Extended Data Fig. 4. We note that the performance of the PRISM-based models in the GDSC test set is correlated with the concordance between the experimentally measured drug's viability profiles in the two screens (Pearson's $R = 0.49$).

PERCEPTION's performance on monotherapy and combinations. The process was carried out in distinct steps as described below:

Step one, quality check to select cell lines, drugs and data points. The Nair dataset comprising monotherapy and combination response was mined from a recent study³⁹ in which the response was measured via the AUC of the dosage–viability curve across eight dosages. As with our GDSC quality check, we only considered drugs with a concordant response profile across the Nair dataset and PRISM (Pearson's Rho > 0.3). AUC values > 1 were removed as they are likely caused by noise in the fitting of the viability curve, owing to noise and higher variability in doses that do not inhibit. This criterion yielded 14 FDA-approved drugs in 21 cell lines, and we focused on them.

Step two, building models. The standard PERCEPTION workflow was used to build a model for these 14 drugs.

Step three, combination response prediction. We extended the prediction to the response to combinations of these 14 drugs studied in this screen (Supplementary Table 5). A combination response in a cell line was predicted by adopting the IDA model across all the single cells from that cell line⁵²; that is, the predicted combination response of n drugs is the effect of the single most effective drug in the combination. Performance was measured using AUC. Throughout our work, the combination response was predicted using the IDA principle.

Step four, performance measurement. As above, we converted the continuous measures of viability to sensitive versus resistant labels. Using these labels, we computed the stratification AUC for monotherapy and combination response prediction.

PERCEPTION's performance on head and neck cancer cell lines. The approach for this prediction was undertaken through specific stages.

Step one, data collection and initial analysis. We obtained the sc-expression data for five head and neck squamous cell cancer PDCs lines, along with their treatment response for eight drugs and combination therapy at two different dosages²⁶. An initial assessment revealed that PERCEPTION was unable to build drug response models with a Spearman correlation greater than 0.3 between predicted and experimental viability using PRISM screens. Therefore, we introduced changes to the PERCEPTION pipeline.

Step two, modification of the PERCEPTION pipeline: building models from GDSC screens. We turned to GDSC screens to build models for drug response, using data from more than approximately 800 cell lines specific to these drugs.

Step three, building models. We considered only the top 3,000 highly expressed genes (with fewer dropouts in the head and neck squamous cell cancer dataset) in common between the bulk expression and PDC datasets to ensure a focused analysis on relevant genes. We then built a PERCEPTION model using these 3,000 genes and GDSC response profiles. The monotherapy and combination responses were calculated following the same methodology used in the GDSC and Nair dataset cases.

Step four, performance measurement. As in the analyses above, we calculated both stratification AUC and correlations for assessing the performance.

Predicting combinations response in patients with multiple myeloma

Response labels, sc-expression of patients' tumors, clustering annotation and mean cluster expression for the multiple myeloma data were mined from the original publication²⁹. No statistical methods were used to pre-determine sample sizes; we used all available samples. We only used the cells annotated as malignant. Predicting the combination response of a patient can be divided into a two-step process: step one, predict the combination response of each clone in that tumor; step two, predict the patient's response from the clone-level combination response. To this end, we first tried to build PERCEPTION response models for the four treatments used in the combination therapy. We succeeded in building PERCEPTION response models for two of the four drugs in the trial that are predictive in cell lines (carfilzomib and lenalidomide; Methods) and used them to predict the treatment response in patients. We first predicted the combination response for each transcriptional cluster (simply referred to here as a clone). To this end, we predicted the response for each of the two drugs separately and computed the killing using the IDA principle; that is, the predicted combination response of n drugs is simply the effect of the single most effective drug in the combination⁵². To overcome the challenge of the discrepancy of dosages used in the clinic versus preclinical testing where our models are built, we z-scale our predicted response profile of a drug across clones; this z-score-predicted response represents the relative response of a clone compared to all others available in the cohort.

In step two, we use this clone-level combination killing profile in a patient to predict the overall patient's response. We considered the predicted response of the least responsive clone found in each patient as that patient's response. This is based on the notion that it would be selected by the treatment and eventually dominate the overall tumor. Performance was measured using AUC. For our model building control, we built random models using either shuffled labels, randomized features in the regression model or a non-predictive model of another drug in the screen for 1,000 times and computed the number of times that the stratification power denoted by AUC is higher than our original model. This proportion is provided as an empirical P value.

Testing prediction strategies for multiple myeloma. Five different strategies were designed to translate clone-level killing into an overall clinical response prediction. These strategies were (1) weighted average response, calculating the average response across all the clones, with each clone's response weighted by its abundance in the tumor; (2) unweighted average response, taking a simple average of the response across all the clones; (3) most-sensitive clone response, using the response of the clone with the highest predicted response; (4) unweighted most-resistant clone response, using the response of the most-resistant clone (the clone with the least response), without weighting; and (5) most-resistant clone response, choosing the response of the most-resistant clone but weighted by its abundance proportion. The performance of these strategies was tested in a cohort comprising responders ($n = 7$) and non-responders ($n = 21$) to the treatment. The accuracy of each strategy was measured using AUC.

Predicting combinations response in breast cancer clinical trial analysis

The pre-filtered 10 \times -based scRNA-seq count data and the cell type annotations of the 65 breast cancer samples (34 patients) were downloaded from GEO (GSE158724). No statistical methods were used to pre-determine sample sizes; we used all available samples. Samples were collected at different time points during the patients' treatments: at the time of screening (S), on day 14 (M) and on day 180 at the end of the trial (E). In our analysis, we considered only the cells annotated as tumor cells. As defined in the primary publication of the dataset⁴⁰, we applied Seurat (v.4.0.5)⁵³. We filtered out samples with fewer than 100 cells. The 38 transcription clusters identified in all 65 samples post-filtering and data processing are as presented in Extended Data Fig. 7a,b. We used the reciprocal principal-component analysis integration workflow to integrate the tumor cells from the remaining samples⁵³. The data were normalized using the SCTransform function, and the top 5,000 variably expressed genes and the first 50 principal components were used in the anchor-based integration step. The first 50 principal components and a k.param value of 20 were used to identify neighbors and the resolution was set to 0.8 to find distinct clusters. We identified 36 different clones, of which only 16 clones were found in the pre-treated samples from patients in arms B and C. The sc-expression of 16 clones was considered in the drug response prediction analysis. The patient response information was obtained from Supplementary Table 12 of the original publication⁴⁰.

We used data from patients with paired samples at time points S and E to study the change in post-treatment response. Extended Data Fig. 7b shows the clonal distribution in each sample processed; all sub-clones (which represent <5% of the cells in the sample) are excluded from our analysis. The default PERCEPTION pipeline was used to build drug response models except for a single change. The top ~2,500 highly expressed genes (ranked by the total number of non-zeroes across all the cancer cells) in the breast cancer dataset that are in common with the cancer cell line bulk expression data were used in the pipeline. The resulting models were used to predict responses at the patient level in a similar manner to what we did for the multiple myeloma data. The controls for the model building were also tested for the breast cancer data, similar to the testing we did for the multiple myeloma data. We note that the number of clusters identified using the standard Seurat pipeline slightly changed when the initial seed for random number generation was changed. These changes did influence the performance by up to 1 in the first significant digit (AUC varied from 0.70 to 0.83 when the seed was changed), but the overall inferences were consistent.

Response models to distinguish responders versus non-responders

We built bulk-based drug response models to compare their performance versus PERCEPTION models in stratifying responders from non-responders in the two clinical trials. To build drug response models

based on bulk expression data, we considered all ~500 cell lines with bulk expression and PRISM-based drug response. For each drug, we randomly divided the data into training (one-third of the cell lines) and test sets (two-thirds of the cell lines). As a feature selection step, we first identified genes whose bulk expression is correlated with the drug viability profile (Pearson's R) in the training set. We considered the correlation for each gene as a measure of information in a gene expression profile and ranked each gene based on the strength of the correlation. While considering the top 100 genes, we built a linear regression model regularized using an elastic net to predict the response to leave-one-out cross-validation, as implemented in R's `glmnet`⁵¹. The resulting model performance was validated on the testing dataset.

To build state-of-the-art bulk-based drug response models as defined in a previous study³³, we generated random-forest-based models in a similar framework as defined above. To make sure that the gene features used in the resulting model predictors are detected to be expressed in the patient sc dataset, we consider genes that overlap in both the cell line bulk expression data and patient sc dataset to build the models. For each drug, we repeated the above model-building steps 100 times and presented the mean and standard error of their performances in stratifying responders from non-responders in their respective clinical trials.

Predicting resistance to tyrosine kinase inhibitors in NSCLC

The sc-expression profiles of 39 biopsies from 25 patients with NSCLC were provided by the original study authors⁴¹. The clinical annotations used for this analysis were mined from the original publication⁴¹ (their Table S1). As in the previous sections, we focused only on the subset of single cells labeled as malignant in the publication. Seurat clustering was performed with resolution = 0.8, dims = 10, number of features = 2,000, scale.factor = 10,000, log-normalization method with minimum cells in a cluster required to be >3 and minimum features required to be >200, to identify a total of 16 clones (Extended Data Fig. 8a). The expression of each transcriptional cluster or clone in a patient is the averaged expression across all single cells associated with that cluster in that given patient and a rank normalization is performed. We successfully built drug response models for dabrafenib, erlotinib, gemcitabine, osimertinib and trametinib. The response observed in the most resistant clone of a patient is considered PERCEPTION's predicted response. We primarily studied the development of drug resistance in the trial. To this end, we defined a term called 'extent of resistance' of a drug, which is the difference between a drug's predicted viability from PERCEPTION and the predicted baseline viability. The predicted baseline viability is defined as the average predicted viability of the respective treatments in all treatment-naïve samples. This difference in response from the naïve state denotes the extent of resistance and is thus named accordingly. We computed both Spearman and Pearson correlations to identify robust correlations.

Literature survey of cross-resistance and cross-sensitivity

To search for evidence available in published papers for a cross-resistant or cross-sensitive drug pair, we used the search term 'drug X AND drug Y' (for example, erlotinib AND gemcitabine) in the PubMed search portal (<https://pubmed.ncbi.nlm.nih.gov>) on 26 December 2021. The resulting clinical trials in the first 50 matches, sorted by best match, were manually curated for outcomes. For preclinical evidence for or against, non-clinical studies testing the combinations were manually curated.

Change of abundance versus predicted resistance of a clone

We first computed and ranked all clones with at least two data points at different time points by their mean predicted resistance across all samples in which they are present. For each clone, we next computed the rate of change of abundance (slope) of the best-fit line of abundance

versus biopsy time from the start of treatment. Finally, we compared this 'rate of change of abundance' with the 'mean predicted resistance' of each clone.

Comparing PERCEPTION's performance on three clinical cohorts

We identified relevant, competing state-of-the-art sc methods for benchmarking PERCEPTION by searching PubMed using the search term 'single-cell expression prediction'. This search yielded only Beyondcell, a state-of-the-art model based on sc-expression²⁷. Additionally, we tested a state-of-the-art bulk model (ATLANTIS) and four alternative methods. The implementation of Beyondcell was downloaded from <https://github.com/cnio-bu/beyondcell> and the default signatures provided by Beyondcell's authors were used. The random-forest-based ATLANTIS was downloaded from <https://github.com/cancerdatasci/atlantis/releases> and used with the default settings. Benchmarking PERCEPTION against these tools to predict patient response, we calculated the AUC for each model in the three clinical cohorts (multiple myeloma, breast cancer and lung cancer). We then calculated the mean AUC across the three cohorts for each model to determine the overall performance.

Testing the most-resistant clone strategy in cell lines

We tested the performance in cell lines for the most-resistant clone strategy to stratify resistant versus sensitive cell lines. To this end, we first clustered the 200 cell lines with Seurat using uniform parameters used across the study, noting 29 clusters and four clusters per cell line. We repeated this process for the head and neck PDCs. The transcriptional cluster or clonal information was obtained from the original publication. We analyzed the sc-expression of primary cells derived from five different patients treated with eight different drugs at two concentrations (Supplementary Table 6), including both monotherapy and combination therapies⁹. We could build PERCEPTION response models for four out of the eight drugs tested (docetaxel, epothilone-b, gefitinib and vorinostat; Pearson's $R > 0.25$). Resistant versus sensitive cells were the top versus bottom 40% cell lines ranked by viability. Our predictions were performed for two dosages \times four monotherapies \times five cell lines. The predicted viability over the 20 (monotherapy, cell line) pairs, comprising four monotherapies \times five cell lines, is correlated with the observed viability, and individual drug-level correlations are provided in Extended Data Fig. 5. We plotted the predicted versus experimental correlations obtained for all data points, and drug levels are provided in Extended Data Figs. 5 and 6.

Drug combinations targeting multiple myeloma clones

To predict combinations for multiple myeloma patients that target multiple clones in the tumor disjointly, and thus have a low likelihood for resistance emergence, we began with all combinations of two drugs with the predictive PERCEPTION model ($n = 44$ drugs) and ranked every pair by a score denoting the extent of their disjoint killing, termed its DKS. This score quantifies the increase in predicted killing compared to the expected (better killing of the two monotherapies) of a drug combination. Out of the 946 possible combinations scanned, 842 pairs showed no improvement over the expected (DKS = 0). Analogously, we next looked for all possible triplets of drug combinations exhaustively ($n = 13,244$). Once validated, this design can be used for creating optimal combinations targeting multiple clones in a patient. Applying this approach to the lung cancer model, we ranked every pair ($n = 946$) by the DKS computed across the four different sc lung cancer cohorts. Out of the 946 possible combinations evaluated, 915 pairs showed no improvement over the monotherapy treatment (DKS = 0). The remaining combinations, with DKS > 0, are shown in Extended Data Fig. 10a,c. We also computed therapy types that are more likely to have high DKS.

Using TrialTrove to test PERCEPTION on predicting response

We reasoned that it would be possible to curate clinical trial data to assess the predictive power of PERCEPTION models in a clinical setting. We used a licensed database, TrialTrove, which has more trials and more detailed and structured information than ClinicalTrials.gov, better facilitating our data extraction⁵⁴. We assembled a collection of clinical trials data of combination therapy, using software we built to parse the TrialTrove database. No statistical methods were used to pre-determine sample sizes.

To test our model, our general approach was to identify combination, multi-arm trials in which one patient arm was administered two drugs, A + B, and another patient arm was administered only drug A. Specifically, we mined trials meeting three criteria: (1) uniform and consistent trial efficacy labels for either median progression-free survival or median overall survival, (2) having at least two arms with treatment design of drug A versus drug A + drug B and (3) targeted therapy treatments (drugs A and B) with predictive PERCEPTION models. Among the several cancer types we investigated, NSCLC was the type for which we could find sufficient and the most abundant homogeneous data for $n > 10$ trials, partly because the median survival data are more readily available for cancers such as NSCLC with poor survival; therefore, we focused on NSCLC in this subsection. An additional filter that was applied is that we must be able to build a PERCEPTION model for both drugs A and B. We next predicted the improvement in response to such combinations over whichever monotherapy was tested in the trial, computed as the response difference between the combination and monotherapy (survival improvement owing to combination), in patients from four NSCLC cohorts with sc-expression^{55–57}, serving as representative samples of sc tumor data of NSCLC patients (total of 18 patients).

We started with a repository of 66,116 oncology trials in phases beyond phase I. To identify combination trials, we used a Python implementation of a modified form of a query suggested by a TrialTrove curator. One of the three fields (Trial Title, Trial Objective, Treatment Plan) should contain any one of the seven strings: ‘combination’, ‘both drugs’, ‘with or without’, ‘combined’, ‘plus’, ‘with and without’, ‘alone or with’ and ‘concurrent’. In addition, the trial keywords must not contain the string ‘single-arm’. To narrow down our list of applicable trials to only those with results of interest, we required (through another Python program) that the trial results field must contain any one of 93 strings such as ORR, OS, PFS, response rate, overall survival, progression-free survival, disease control rate, and so on. To identify trials that used drugs that PERCEPTION can model, we processed the fields named ‘primary tested drug’ and ‘other tested drug’ to require that together these two fields must contain at least two drugs that can be modeled by PERCEPTION and at least one drug that is a targeted therapy rather than chemotherapy. The ‘primary tested drug’ and ‘other tested drug’ fields are already normalized for synonyms. The overall goal of the three filters (combination trials, results available, PERCEPTION-suitable drugs used) was to eliminate false negatives, trials that would not be useful in testing PERCEPTION-built models. Trials that survived these three filters were then curated manually to obtain accurate arm information and results.

We measured our performance by computing a correlation between PERCEPTION’s predicted improvement to combination versus survival improvement owing to the combination observed in the respective clinical trials. Separate analyses for overall and progression-free survival were also done. However, we note the small cohorts available for these two analyses. Repeating this analysis in a drug-specific manner, we focused on trials of different drug combinations ($n = 6$) with erlotinib, the only targeted therapy with a sufficient number of unique combination trials.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The entire collection of the processed datasets used in this manuscript, including preclinical models of cancer cell lines and PDCs, can be accessed in the Zenodo repository (<https://zenodo.org/record/7860559>)⁵⁸. We collected the bulk-expression and drug response profiles generated in cancer cell lines curated from the DepMap portal (<https://depmap.org/portal/download>) (version 20Q1). The sc-expression of 205 cancer cell lines was generated in a previous study³⁴ and was downloaded from https://singlecell.broadinstitute.org/single_cell/study/SCP542/pan-cancer-cell-line-heterogeneity#study-download. The sc-expression profiles of patients with multiple myeloma were downloaded from the original study (their supplementary Table 2; https://static-content.springer.com/esm/art%3A10.1038%2Fs41591-021-01232-w/MediaObjects/41591_2021_1232_MOESM3_ESM.xlsx); data from patients with breast cancer were downloaded from GEO ([GSE158724](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158724)) and data from patients with NSCLC were provided by the original study authors⁴¹.

Code availability

The scripts to replicate each step of results and plots can be accessed in a GitHub repository (https://github.com/ruppinlab/SCPO_submission). We used open-source R versions 4.0 through 4.2 to generate the figures. Wherever required, commercially available Adobe Illustrator was used to create the figure grids.

References

1. Tsimberidou, A. M., Fountzilas, E., Nikanjam, M. & Kurzrock, R. Review of precision cancer medicine: evolution of the treatment paradigm. *Cancer Treat. Rev.* **86**, 102019 (2020).
2. Huang, K., Xiao, C., Glass, L. M. & Critchlow, C. M. Machine learning applications for therapeutic tasks with genomics data. *Patterns* **2**, 100328 (2021).
3. Bhinder, B., Gilvary, C., Madhukar, N. S. & Elemento, O. Artificial intelligence in cancer research and precision medicine. *Cancer Discov.* **11**, 900–915 (2021).
4. Singla, N. & Singla, S. Harnessing big data with machine learning in precision oncology. *Kidney Cancer J.* **18**, 83–84 (2020).
5. Senft, D., Leiserson, M. D. M., Ruppin, E. & Ronai, Z. Precision oncology: the road ahead. *Trends Mol. Med.* **23**, 874–898 (2017).
6. Tsimberidou, A. M., Fountzilas, E., Bleris, L. & Kurzrock, R. Transcriptomics and solid tumors: the next frontier in precision cancer medicine. *Semin. Cancer Biol.* **84**, 50–59 (2022).
7. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).
8. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88 (2019).
9. Sawabata, N. Circulating tumor cells: from the laboratory to the cancer clinic. *Cancers* **12**, 3065 (2020).
10. Beaubier, N. et al. Integrated genomic profiling expands clinical options for patients with cancer. *Nat. Biotechnol.* **37**, 1351–1360 (2019).
11. Hayashi, A. et al. A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. *Nat. Cancer* **1**, 59–74 (2020).
12. Rodon, J. et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat. Med.* **25**, 751–758 (2019).
13. Tanioka, M. et al. Integrated analysis of RNA and DNA from the phase III trial CALGB 40601 identifies predictors of response to trastuzumab-based neoadjuvant chemotherapy in HER2-positive breast cancer. *Clin. Cancer Res.* **24**, 5292–5304 (2018).

14. Vaske, O. M. et al. Comparative tumor RNA sequencing analysis for difficult-to-treat pediatric and young adult patients with cancer. *JAMA Netw. Open* **2**, e1913968 (2019).
15. Wong, M. et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nat. Med.* **26**, 1742–1753 (2020).
16. Lee, J. S. et al. Synthetic lethality-mediated precision oncology via the tumor transcriptome. *Cell* **184**, 2487–2502 (2021).
17. Dinstag, G. et al. Clinically oriented prediction of patient response to targeted and immunotherapies from the tumor transcriptome. *Med* **4**, 15–30.e8 (2023).
18. Castro, L. N. G., Tirosch, I. & Suà, M. L. Decoding cancer biology one cell at a time. *Cancer Discov.* **11**, 960–970 (2021).
19. Wensink, G. E. et al. Patient-derived organoids as a predictive biomarker for treatment response in cancer patients. *npj Precis. Oncol.* **5**, 30 (2021).
20. Yao, Y. et al. Patient-derived organoids predict chemoradiation responses of locally advanced rectal cancer. *Cell Stem Cell* **26**, 17–26 (2020).
21. de Witte, C. J. et al. Patient-derived ovarian cancer organoids mimic clinical response and exhibit heterogeneous inter-and intrapatient drug responses. *Cell Rep.* **31**, 107762 (2020).
22. Shalek, A. K. & Benson, M. Single-cell analyses to tailor treatments. *Sci. Transl. Med.* **9**, eaan4730 (2017).
23. Adam, G. et al. Machine learning approaches to drug response prediction: challenges and recent progress. *npj Precis. Oncol.* **4**, 19 (2020).
24. Zhu, S. et al. Advances in single-cell RNA sequencing and its applications in cancer research. *Oncotarget* **8**, 53763–53779 (2017).
25. Kim, K. T. et al. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* **17**, 80 (2016).
26. Suphavilai, C. et al. Predicting heterogeneity in clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. *Genome Med.* **13**, 189 (2021).
27. Fustero-Torre, C. et al. Beyondcell: targeting cancer therapeutic heterogeneity in single-cell RNA-seq data. *Genome Med.* **13**, 187 (2021).
28. Ianevski, A. et al. Patient-tailored design for selective co-inhibition of leukemic cell subpopulations. *Sci. Adv.* **7**, eab4038 (2021).
29. Cohen, Y. C. et al. Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing. *Nat. Med.* **27**, 491–503 (2021).
30. Ledergor, G. et al. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat. Med.* **24**, 1867–1876 (2018).
31. Sade-Feldman, M. et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998–1013 (2018).
32. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
33. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576 (2017).
34. Kinker, G. S. et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).
35. Plana, D., Palmer, A. C. & Sorger, P. K. Independent drug action in combination therapy: implications for precision oncology. *Cancer Discov.* **12**, 606–624 (2022).
36. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucl. Acids Res.* **41**, D955–D961 (2012).
37. Seashore-Ludlow, B. et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* **5**, 1210–1223 (2015).
38. Corsello, S. M. et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **1**, 235–248 (2020).
39. Nair, N. U. et al. A landscape of response to drug combinations in non-small-cell lung cancer. *Nat. Commun.* **14**, 3830 (2023).
40. Griffiths, J. I. et al. Serial single-cell genomics reveals convergent subclonal evolution of resistance as patients with early-stage breast cancer progress on endocrine plus CDK4/6 therapy. *Nat. Cancer* **2**, 658–671 (2021).
41. Maynard, A. et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* **182**, 1232–1251 (2020).
42. Noronha, A. et al. AXL and error-prone DNA replication confer drug resistance and offer strategies to treat EGFR-mutant lung cancer. *Cancer Discov.* **12**, 2666–2683 (2022).
43. Pluchino, K. M., Hall, M. D., Goldsborough, A. S., Callaghan, R. & Gottesman, M. M. Collateral sensitivity as a strategy against cancer multidrug resistance. *Drug Resist. Updat.* **15**, 98–105 (2012).
44. Bartholomeusz, C. et al. Gemcitabine overcomes erlotinib resistance in EGFR-overexpressing cancer cells through downregulation of Akt. *J. Cancer* **2**, 435–442 (2011).
45. Moore, M. J. et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J. Clin. Oncol.* **25**, 1960–1966 (2007).
46. Shin, S., Park, C. M., Kwon, H. & Lee, K.-H. Erlotinib plus gemcitabine versus gemcitabine for pancreatic cancer: real-world analysis of Korean national database. *BMC Cancer* **16**, 443 (2016).
47. Luo, J. et al. Erlotinib and trametinib in patients with EGFR-mutant lung adenocarcinoma and acquired resistance to a prior tyrosine kinase inhibitor. *JCO Precis. Oncol.* **5**, 55–64 (2021).
48. Mariotto, A. B. et al. Projections of the cost of cancer care in the United States: 2010–2020. *J. Natl. Cancer Inst.* **103**, 117–128 (2011).
49. Svensson, V. Droplet scRNA-seq is not zero-inflated. *Nat. Biotech.* **38**, 147–150 (2020).
50. Cao, Y., Kitanovski, S., Küppers, R. & Hoffmann, D. UMI or not UMI, that is the question for scRNA-seq zero-inflation. *Nat. Biotechnol.* **39**, 158–159 (2021).
51. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Soft.* **33**, 1–22 (2010).
52. Ling, A. & Huang, R. S. Computationally predicting clinical drug combination efficacy with cancer cell line screens and independent drug action. *Nat. Commun.* **11**, 5848 (2020).
53. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
54. Stergiopoulos, S., Getz, K. A. & Blazynski, C. Evaluating the completeness of ClinicalTrials.gov. *Ther. Innov. Regul. Sci.* **53**, 307–317 (2019).
55. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
56. Song, Q. et al. Dissecting intratumoral myeloid cell plasticity by single cell RNA-seq. *Cancer Med.* **8**, 3072–3085 (2019).
57. Zilionis, R. et al. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* **50**, 1317–1334 (2019).
58. Sinha, S. Predicting patient treatment response and resistance via single-cell transcriptomics of their tumors (O1) [Data set]. Zenodo <https://doi.org/10.5281/zenodo.7860559> (2022).

Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute (NCI), NIH grants R01CA231300 (T.G.B.), R01CA204302

(T.G.B.), R01CA211052 (T.G.B.), R01CA169338 (T.G.B.) and U54CA224081 (T.G.B.). This work used the computational resources of the NIH High-Performance Computing Biowulf cluster (<http://hpc.nih.gov>). We acknowledge and thank the NCI for providing financial and infrastructural support. Thanks to K. Wang, S. Rajagopal and Z. Ronai for their valuable feedback and discussion. Special thanks to J. I. Griffiths and A. H. Bild for clarifying the patient response data in reference 40 and for their helpful feedback.

Author contributions

S.S., R.V., A.A.S. and E.R. conceived the framework of the analysis. E.R. and A.A.S. mentored and guided the study. S.S. and R.V. led the analysis of the development of the models and most of the testing. A.A.S., A.V.K., R.V. and S.S. performed the analysis related to clinical trials curation and data analysis. A.A.S., S.M., S.R.D., N.U.N., M.G.J. and N.Y. worked on the revisions for model validation and further testing and development of the software. W.W., D.L.K., C.M.B. and T.G.B. provided the lung cancer data and aided in its analysis. O.V.S., I.G., K.D.A., C.M.B. and C.J.T. contributed to finding relevant dosages to translate *in vitro* to *in vivo* results. S.S., R.V., A.A.S., E.R., P.J., C.H.B. and T.G.B. wrote the initial draft of the manuscript; S.S., S.M., A.A.S. and E.R. carried out the revisions.

Competing interests

S.S., R.V., A.A.S. and E.R. are inventors on a provisional patent application covering the methods in PERCEPTION. E.R. is a co-founder of Medaware, Metabomed and Pangea Biomed (divested from the latter). E.R. serves as a non-paid scientific consultant to Pangea Biomed, a company developing a precision oncology SL-based multi-omics approach, with emphasis on bulk tumor transcriptomics.

T.G.B. is an advisor to Array/Pfizer, Revolution Medicines, Springworks, Jazz Pharmaceuticals, Relay Therapeutics, Rain Therapeutics and Engine Biosciences, and receives research funding from Novartis, Strategia, Kinnate and Revolution Medicines. The work in the laboratory of C.H.B. was funded in part by Amgen and Novartis. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at
<https://doi.org/10.1038/s43018-024-00756-7>.

Supplementary information The online version contains supplementary material available at
<https://doi.org/10.1038/s43018-024-00756-7>.

Correspondence and requests for materials should be addressed to Sanju Sinha or Eytan Ruppin.

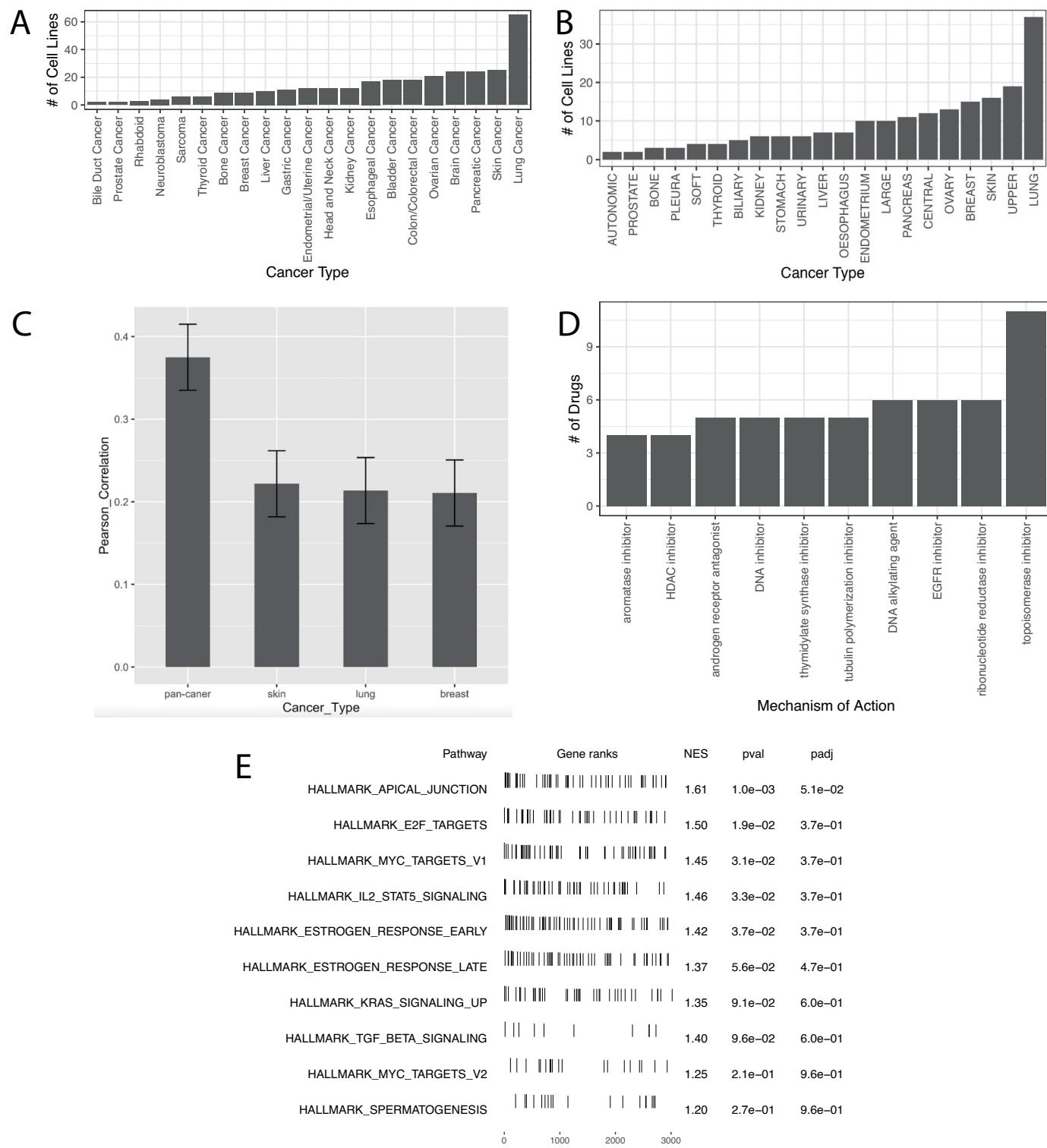
Peer review information *Nature Cancer* thanks Federica Eduati and Tuomas Tammela for their contribution to the peer review of this work.

Reprints and permissions information is available at
www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024

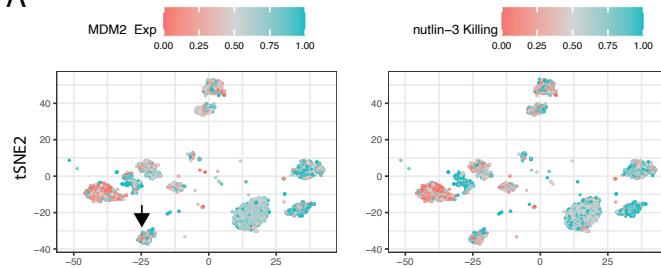
¹Cancer Data Science Laboratory, National Cancer Institute, Bethesda, MD, USA. ²University of Maryland, College Park, MD, USA. ³Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. ⁴Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA. ⁵Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA, USA. ⁶Integrative Program in Quantitative Biology, University of California, San Francisco, San Francisco, CA, USA. ⁷Whitehead Institute, Cambridge, MA, USA. ⁸Rancho BioSciences, San Diego, CA, USA. ⁹Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA. ¹⁰Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. ¹¹Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA. ¹²Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ¹³Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. ¹⁴Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. ¹⁵Chan Zuckerberg Biohub Investigator, San Francisco, CA, USA. ¹⁶Present address: NCI-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, San Diego, CA, USA. ¹⁷These authors contributed equally: Sanju Sinha, Rahulsimham Vigesna. ✉e-mail: ssinha@sbpdiscovery.org; eytan.ruppin@nih.gov



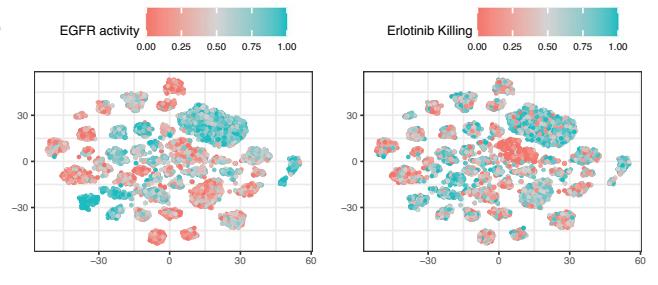
Extended Data Fig. 1 | Overview of PERCEPTION model's training data and features. **A)** Cancer type distribution of the 318 cell lines used during the bulk expression training of PERCEPTION (step 1). **B)** Similarly, showing the cancer type distribution of the 169 cell lines used during the sc-expression training of PERCEPTION (step 2) **C)** The performance of PERCEPTION in predicting response in unseen cell lines when built via (1) pan-cancer models: all available cell lines ($N = 169$) are used for training the model, (2) Cancer-type specific: trained only on cell lines of the same cancer type as those used in the testing ($N = 16$ melanoma cell lines, 37 lung cancer cell lines and 15 breast cancer cell lines, as we used the PERCEPTION to predict the patient's treatment response in three clinical

trial cohorts from skin, lung, and breast cancer, we compared the pan-cancer model with these three individual cancer-type models). No statistical test was performed to compare groups. Error bars indicate the standard error of the mean (SEM), reflecting data variability. **D)** Major classes of mechanism of action of the 133 FDA-approved drugs that were studied here. No statistical test was performed to compare between groups. **E)** Top pathways enriched in frequently appearing features/genes in the PERCEPTION models. This is computed using a GSEA rank test across all hallmark pathways. To assess the statistical significance of these scores, a permutation test was performed.

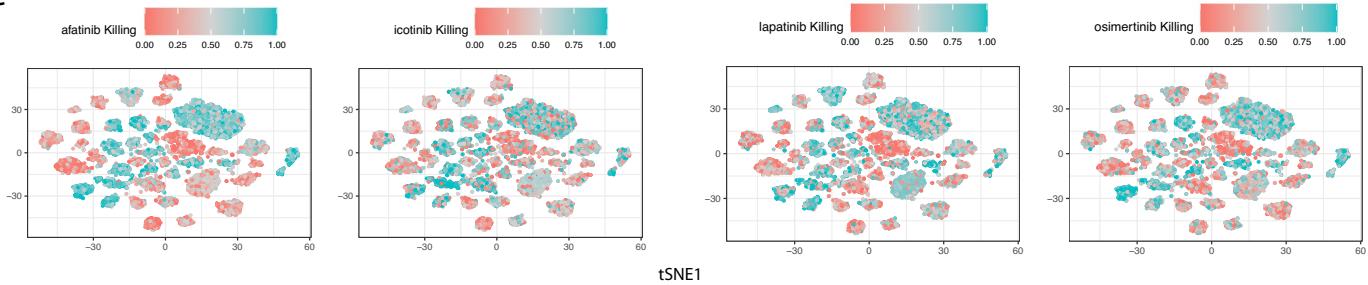
A



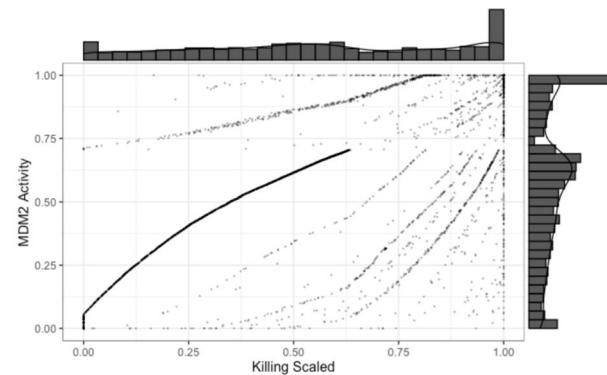
B



C

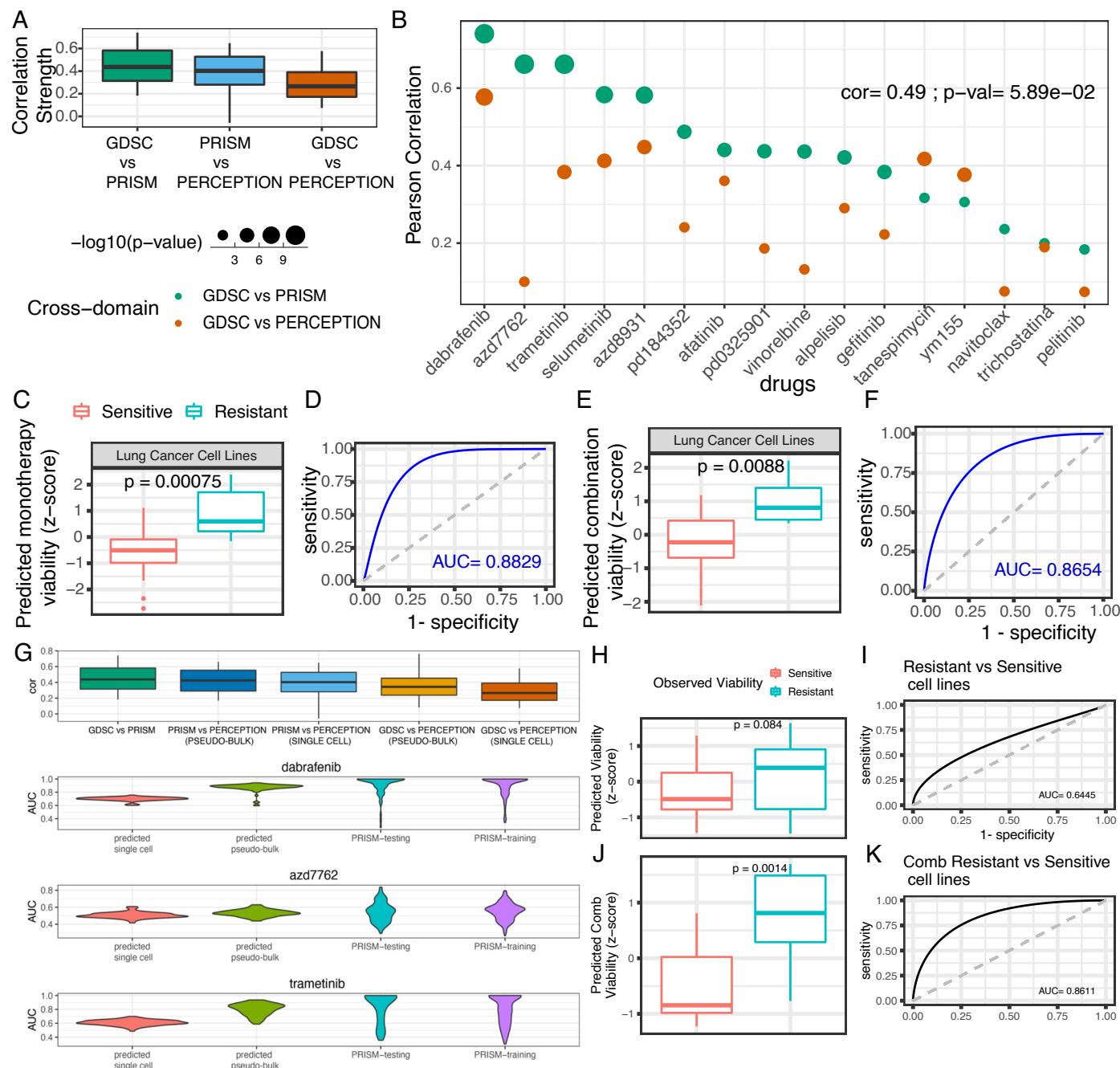


D



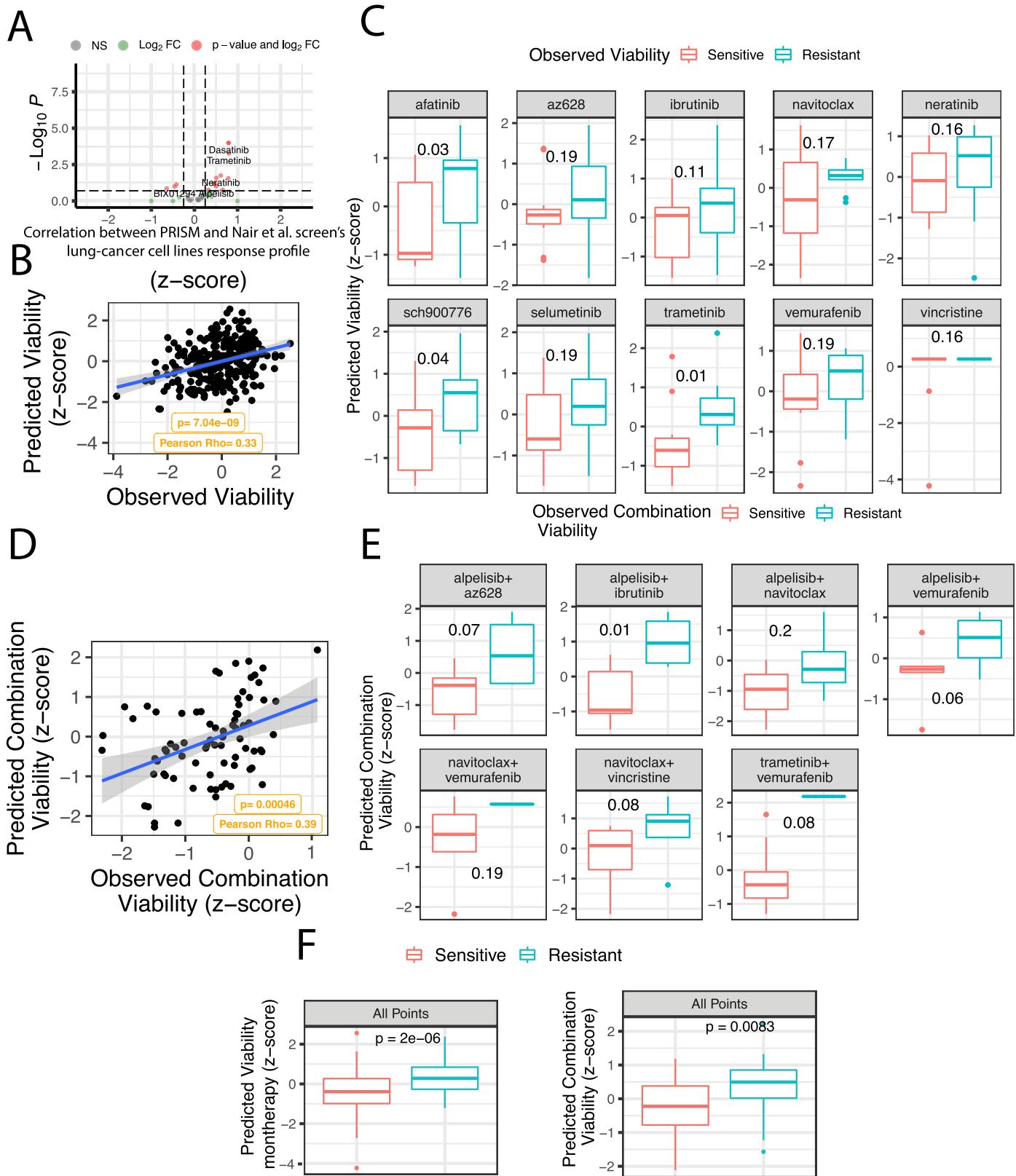
Extended Data Fig. 2 | Visualization of PERCEPTION's ability to predict viability at four recent EGFR inhibitors vs the EGFR pathway activity at single-cell resolution. **A)** The top-most panel visualizes the PERCEPTION predicted killing by nutlin-3, a canonical MDM2 antagonist and the expression of MDM2 for every single cell (each point) in the top and bottom tSNE plot, respectively. The intensity of the color denotes the extent of predicted killing in the right panel and measured MDM2 expression in the left panel. 3566 single-cells from nine p53 WT lung cancer cell lines are depicted. The tSNE clustering is performed using the expression of all the genes. **B)** A similar display visualizes PERCEPTION's predicted killing and the EGFR pathway signature expression

across 12,482 individual lung cancer cells. **C)** The four panels visualize predicted killing by four EGFR inhibitors, afatinib, icotinib, lapatinib, osimertinib, in every single cell (each point) via a tSNE plot, respectively. Here, the color of each point denotes the extent of predicted killing. In this figure, we provide data on 12,482 individual lung cancer cells. The tSNE clustering is performed using the expression profiles of all the genes. **D)** We present here the correlation between the predicted killing effect of nutlin-3 from the PERCEPTION prediction of each cell (x-axis) and the MDM2 gene expression in that single cell, where they are found to be strongly correlated. "MDM2 Activity" on the y-axis denotes MDM2 gene expression.



Extended Data Fig. 3 | Evaluating PERCEPTION's Efficacy in Unseen Lung Cancer Cell Line Screens. **A)** Correlation Analysis: Examines the relationships across three platforms - "GDSC vs. PRISM", "PRISM vs. PERCEPTION" (cross-validation), and "GDSC vs. PERCEPTION". Drug response predictions at single-cell resolution were aggregated to represent overall cell line responses. **B)** These cross-platform correlations are provided at a drug level. Significance of correlations assessed using Pearson's r test. **C)** Monotherapy Predictions by PERCEPTION: Showcases the predicted viability of monotherapies based on cell line-specific sc-expression, comparing resistant ($N = 72$) and sensitive ($N = 84$) lines using boxplots. Significance determined by one-tailed Wilcoxon rank-sum test. **D)** Sensitivity-Specificity Analysis: The receiver operator curve illustrates the balance between sensitivity and specificity in distinguishing between sensitive and resistant cell lines. Area under the curve (AUC) values are noted, with the dashed line representing random-model performance. **E) & F)**

Drug Combination Response Predictions: Depict PERCEPTION's predictions for drug combination responses in resistant ($N = 28$) vs. sensitive ($N = 24$) cell lines. **G)** Single-cell vs. Pseudo-bulk Level Analysis in PRISM Screens: Extends the analysis in panel A to single-cell and pseudo-bulk levels, highlighting the improved performance in pseudo-bulk data. The comparison includes predicted AUC values at both levels and experimental AUC values in PRISM for dabrafenib, AZD-7762, and trametinib, covering both testing ($N = 80$) and training cell lines ($N = 318$). **H-K)** Patient-Derived H&N Primary Cell Analysis: **H)** Prediction of Monotherapy Response: PERCEPTION's predicted viability in resistant ($n = 16$) vs. sensitive ($n = 16$) lines. **I)** ROC Curve Analysis: Illustrates model's prediction capability (sensitivity and specificity) for resistant vs. sensitive lines. AUC values are presented. **J) & K)** Combination Treatment Response: Similar analysis for combination treatments, comparing resistant (12) to sensitive (12) lines. All box plots show median, 25th/75th percentiles, and range.



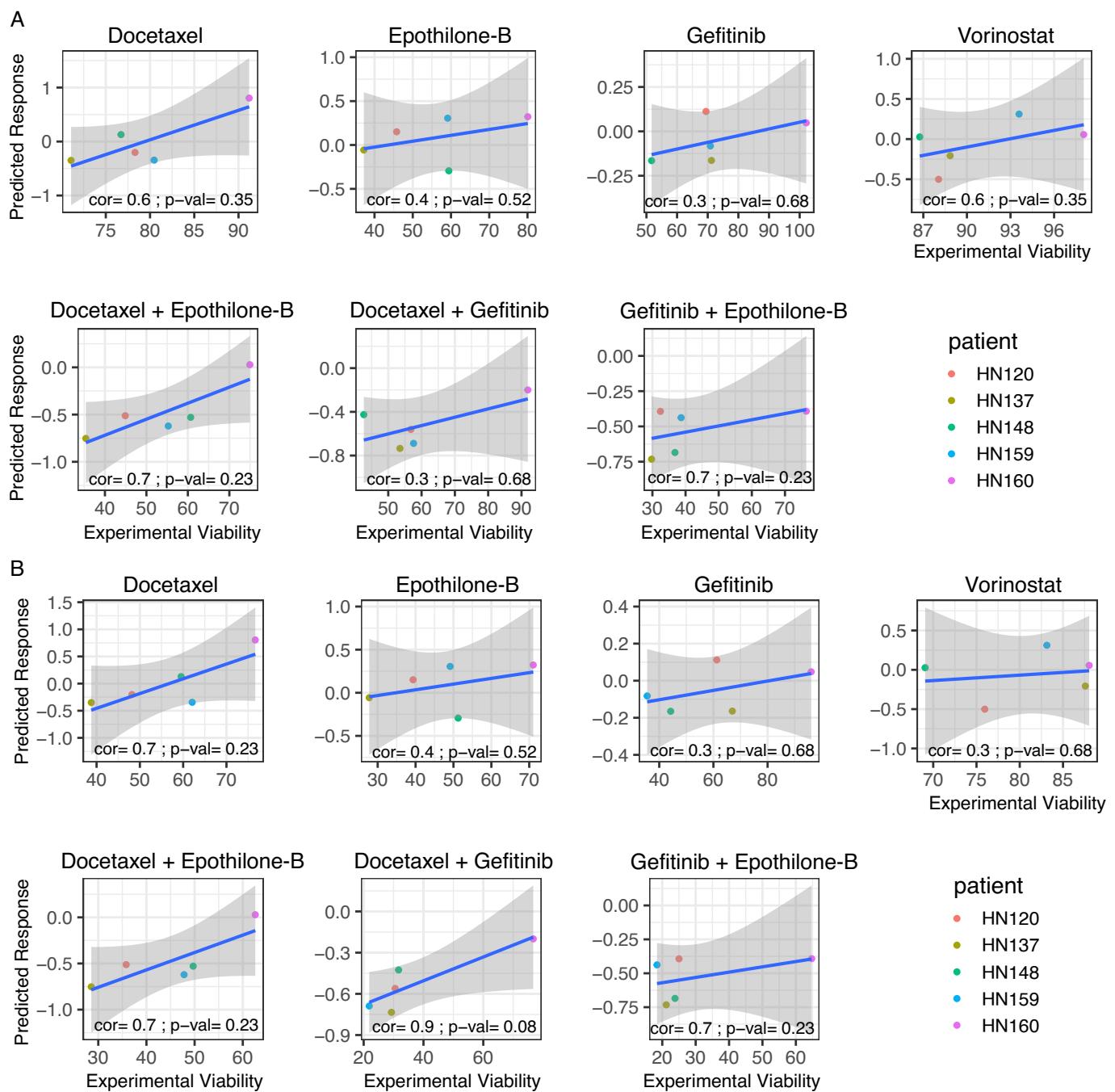
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Quality Control and Predictive Analyses in Lung

Cancer Cell Line Screens. **A)** Concordance between Lung Cancer and PRISM Screens: Illustrates the correlation (Rho on x-axis) and significance (y-axis) between our lung cancer screen and PRISM. Focuses on cell lines showing significantly positive correlation, as indicated by Pearson's r test p-value. **B)** Predicted vs. Observed Viability Comparison: Analyzes the correlation between predicted and observed cell viability ($N = 94$ viability observations each, both centered and scaled). Pearson correlation and significance are noted. A best fit line with a 95% confidence interval is shown. **C)** Viability Prediction in Top vs. Bottom 50% Cell Lines: Compares predicted viability in resistant ($N = 11$, bottom 50%) versus sensitive ($N = 10$, top 50%) cell lines for each drug. Uses one-tailed Wilcoxon rank-sum test for statistical significance, presented for

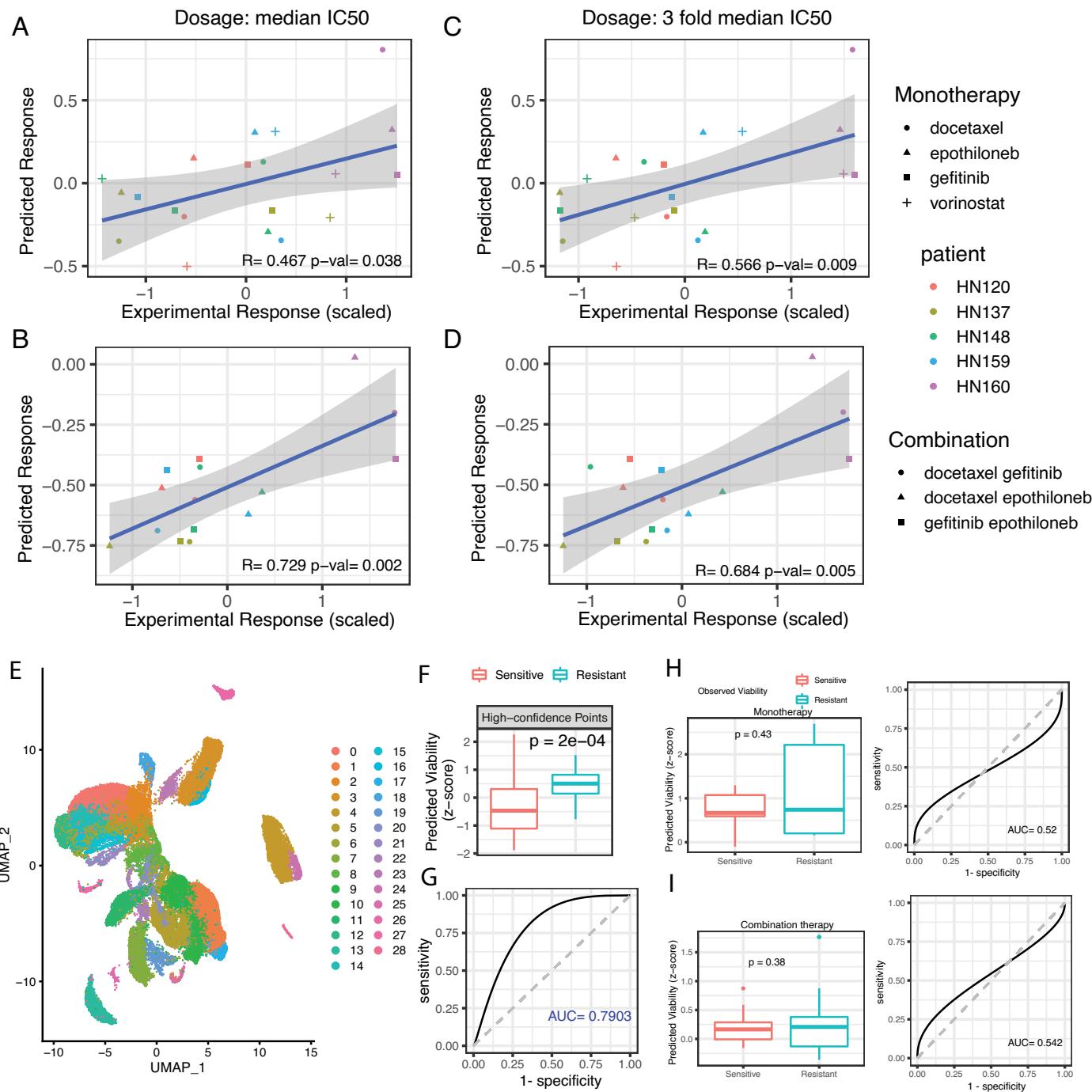
each drug. **D)** Combination Response Prediction in 21 Lung Cancer Cell Lines:

Similar to panel B, this compares predicted versus observed combination viability ($N = 49$ viability observations each), with Pearson correlation and significance provided. A best fit line with a 95% confidence interval is included. **E)** Combination Viability Prediction in Top vs. Bottom 50% Cell Lines: Analyzes predicted combination viability (centered and scaled) for resistant ($N = 11$) and sensitive ($N = 10$) cell lines (based on observed viability) across 7 drug pairs. Uses one-tailed Wilcoxon rank-sum test for significance, presented for each combination. **F)** Consolidated Analysis of Monotherapies and Combinations: Integrates data from distinct drugs in panel E for combined analysis of monotherapies ($N = 188$) and drug combinations ($N = 98$). All box plots show median, 25th/75th percentiles, and range.



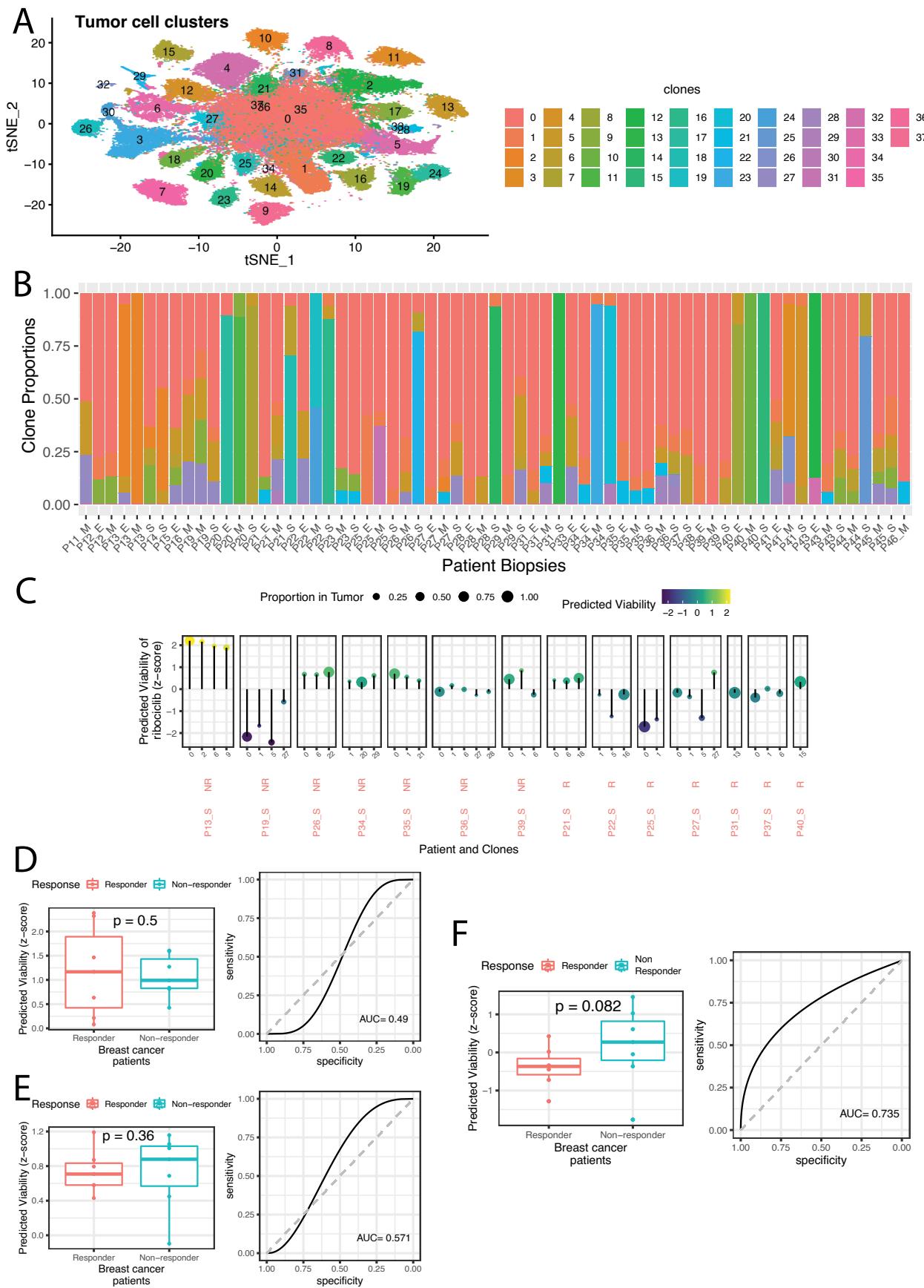
Extended Data Fig. 5 | The predicted vs. experimental correlations obtained for individual treatments. Each scatter plot compares the experimentally observed cell viability (x-axis; at median IC₅₀ concentration) to the predicted viability (y-axis; rescaled AUC value) for the four drugs docetaxel, epothilone-b, gefitinib, and vorinostat (top four) and the pairwise combinations among {docetaxel, epothilone-b, gefitinib} (bottom three). Each dot represents the

response of patient-derived cell lines (N = 5, color coded) for the drugs they were screened with. The Spearman rank correlation (cor) is provided at the bottom of each plot. These plots are provided for the following treatment concentrations - A) median IC₅₀ B) one-third of median IC₅₀. The error bands in all panels of this figure show 95% confidence interval of the fit.



Extended Data Fig. 6 | Correlation of Predicted and Observed Viability in Monotherapies and Combination Treatments in Cell Lines. Each scatter plot compares experimental cell viability ($N = 20$, x-axis; scaled per drug treatment) with predicted viability ($N = 20$, y-axis; rescaled AUC value). Points represent patient-derived cell line responses, color-coded by line and shape-coded by drug. Pearson correlation (R) is noted in each plot's lower right corner. All panels feature error bands showing the 95% confidence interval of the fit. **A)** Monotherapy Response at Median IC50: Relation between monotherapy response and experimental response ($N = 20$ each). **B)** Combination Therapy Response at Median IC50: Similar analysis for combination therapy ($N = 15$ each). **C)** Monotherapy Response at 3x Median IC50: Examines monotherapy response at higher concentration ($N = 20$ each). **D)** Combination Therapy at 3x Median IC50: Analyzes combination therapy response at increased concentration ($N = 15$ each). **E-G)** Monotherapy and Combination Response Prediction in Lung Cancer Cell Lines: **E)** UMAP Clustering: Represents 53,514 cells from 199 cell lines (-300 cells/line) using sc-expression, identifying 29

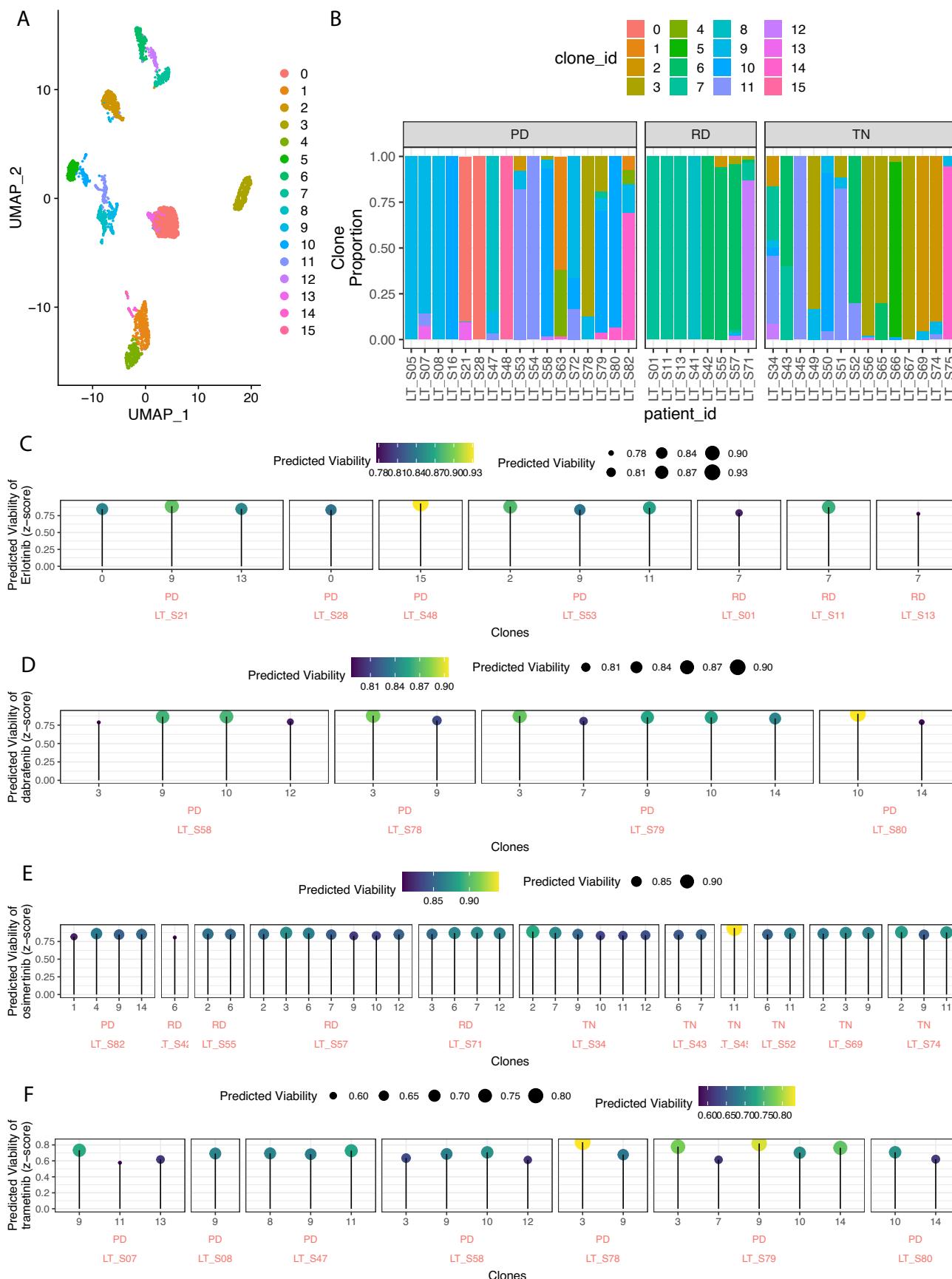
clusters with cells from four unique sub-clones. **F-G)** Predicted Viability Based on Most-Resistant Clone: Viability predictions for 21 lung cancer cell lines ($N = 11$ resistant & 10 sensitive cell lines), considering the most resistant clone. Statistical significance assessed with two-sided Wilcoxon rank-sum test. **H-I)** Monotherapy and Combination Response Prediction in Patient-Derived HNSC Primary Cells ($N = 5$): **H)** Monotherapy Response Based on Most-Resistant Clone: Presents PERCEPTION predicted viability and resistance vs. sensitivity stratification ($N = 2$ resistant & 3 sensitive). Includes drugs docetaxel, epothilone-b, gefitinib, and vorinostat. **I)** Combination Response: Similar analysis for combination treatments. Both panels include a left-side plot for predicted viability in resistant ($N = 2$) vs. sensitive ($N = 3$) lines and a right-side ROC plot showing prediction power (sensitivity and specificity). AUC values are provided, with the dashed line indicating random-model performance. Statistical analysis performed with two-sided Wilcoxon rank-sum test. All box plots depict median, 25th/75th percentiles, and range.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Comparing PERCEPTION with Existing Bulk Response Models in a Breast Cancer Clinical Trial. **A)** tSNE Transcriptional Clustering: Displays 36 transcriptional tumor clusters identified in the trial, integrating cells from 34 patients at three time points. Clusters, color-coded and defined in the legend, were derived using Seurat package. **B)** Malignant Sub-Clone Abundance: Shows the distribution of malignant sub-clones (y-axis) in breast cancer samples (x-axis), based on sc-expression. Different sub-clones are color-coded in the legend. Sample labels on the x-axis indicate patient id and time point of collection (“_S” - day 0, “_M” - day 14, “_E” - day 180). **C)** Pre-Treatment Clone-Level Response in Arms B and C: Predicted ribociclib viability (y-axis) versus various clones in pre-treatment samples (x-axis). Response status is displayed at the top of each column, with sample names below. Dot sizes represent the proportion of each cluster/clone, with a color scale indicating predicted viability (dark blue for low, yellow for high). **D-E)** Stratification Power of PERCEPTION vs Published

Models: **D)** Bulk Expression-Based Models: Compares PERCEPTION with models trained only on bulk expression (N = 7 responders and 7 non-responders). **E)** Models Not Tuned on sc-Expression: PERCEPTION compared against models without sc-expression tuning (N = 7 responders and 7 non-responders). Both panels include deterministic model generation (seed=1) for training and test sets. Left-side plots present PERCEPTION predicted viability in responders vs. non-responders. Right-side ROC plots depict prediction power (sensitivity and specificity), with AUC values near the lower right corner. The dashed diagonal line indicates performance of a random model. Statistical significance assessed using two-sided Wilcoxon rank-sum test. **F)** Stratification Using Average sc-Viability: Stratifies responders (N = 7) vs. non-responders (N = 7) in combination therapy arms using average sc-viability in the FELINE trial. Statistical significance evaluated by two-sided Wilcoxon rank-sum test. Box plots show median, 25th/75th percentiles, and range.

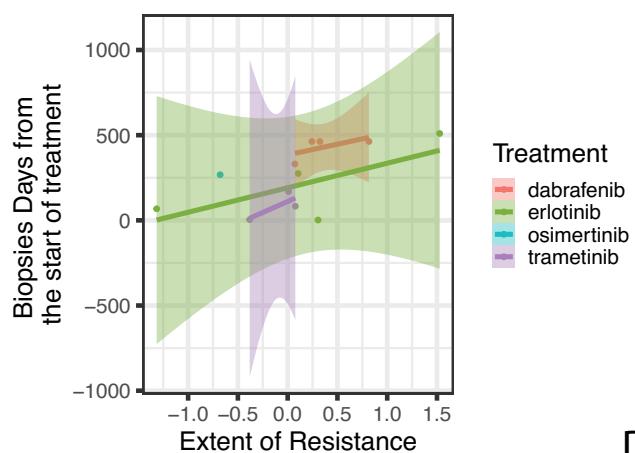


Extended Data Fig. 8 | See next page for caption.

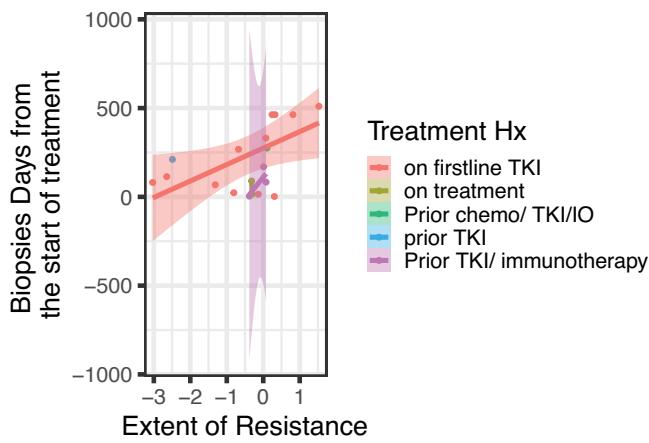
Extended Data Fig. 8 | Pre-processing and predicting clone level response in lung cancer patient cohort. (A) A UMAP of 3671 malignant cells derived from 25 patients with 26,485 genes are clustered using Seurat considering the first 10 axes with the most variance. Each clone (a transcriptional cluster) output is annotated using a color where the legend is provided on the right. (B) The proportions of

these clones (y-axis) are provided in each patient (x-axis) faceted by the time point at which these biopsies are collected. (C-F) Predicted viability of the four tyrosine kinase inhibitors: erlotinib, dabrafenib, osimertinib, and trametinib, in respective order, is provided at a clonal level for each patient where response status is provided at the bottom of each facet.

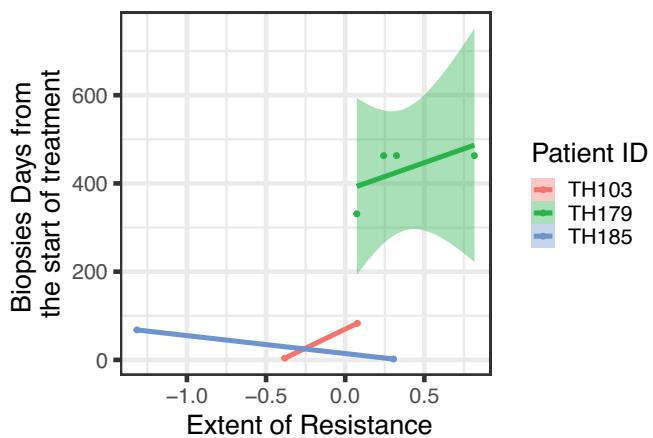
A



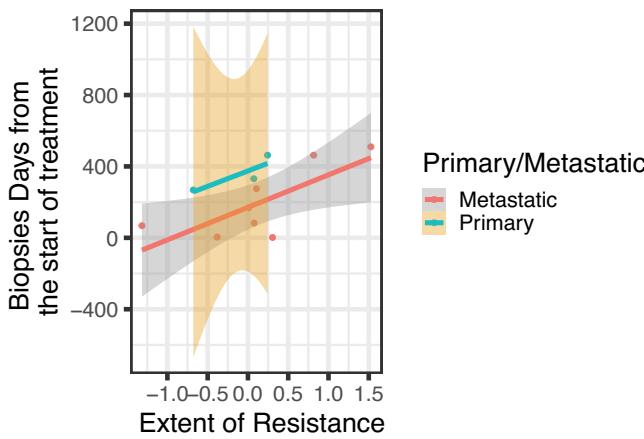
B



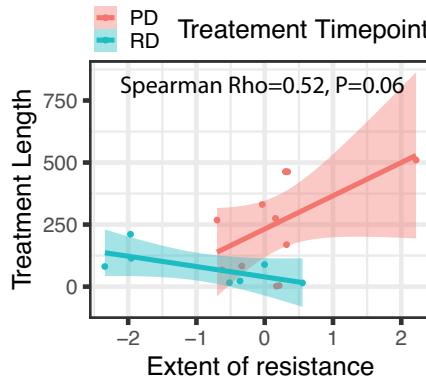
C



D

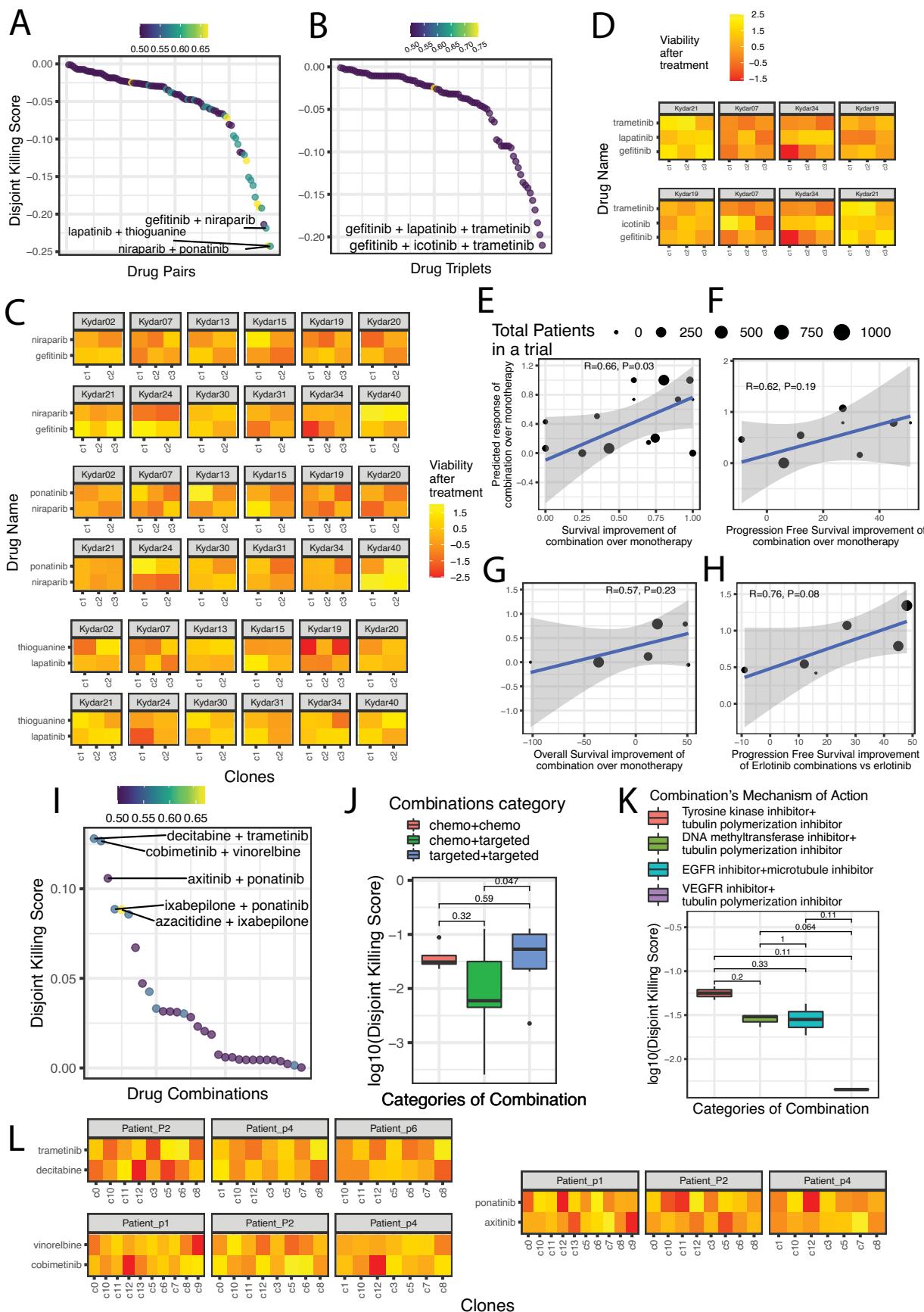


E



Extended Data Fig. 9 | Correlation between the elapsed treatment time and estimated resistance holds true across different conditions. In A-D, The extent of resistance to a treatment from the baseline (x-axis) is correlated with the treatment elapsed time (Number of days from the start of the treatment before the biopsy was taken) (y-axis). (A) The points and line colors denote the treatment administered to the patients listed by the right legend. (B) Color denotes prior treatment. (C) Color denotes the patient's ID. (D) Color denotes

whether the disease is metastatic or primary at the time of biopsy. (E) Extent of Resistance was calculated using bulk-expression of the tumor, where the increase with "Treatment Elapsed time" is positive, however, insignificant, and weaker than when the patient response is taken as the most-resistant clone available response. The error bands in all panels of this figure show 95% confidence interval of the fit.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Identifying Optimal Drug Combinations for Multiple Myeloma and Lung Cancer Patients. **A)** Median Disjoint Killing Score (DKS) in Myeloma: For 94 drug pairs with positive DKS, the median DKS (y-axis) is plotted against each pair (x-axis). Color intensity denotes the proportion of patients ($N = 12$) with DKS > 0 , with the top pairs labeled. Legend for color intensity is at the top. **B)** DKS for Triplets: Similar analysis for drug triplets. **C)** Clone-Level Disjoint Killing for Top Pairs: Viability profiles of clones for top pairs from C are shown for each patient (facet), with color intensity indicating post-treatment viability of each clone (x-axis) for a given drug (y-axis). Legend on the right. **D)** Clone-Level Disjoint Killing for Triplets: Analogous to C, but for drug triplets ($N = 86$, Triplets with DKS > 0). **E-L)** Analysis in Lung Cancer: **E)** Correlation in Clinical Trials: Examines the correlation between response difference of combination vs monotherapy (x-axis) and observed survival difference in combination vs single-treatment arms. Dot size represents patient numbers,

with a best-fit line shown. Legend for dot sizes and error bands showing 95% confidence interval are at the top. Weighted Pearson's r test p-value denotes correlation significance. **F-H)** Repeated for progression-free survival, overall survival, and erlotinib combinations. **I)** DKS for Lung Cancer Drug Combinations: Median DKS (y-axis) for 31 positive pairs plotted against each pair (x-axis). Color intensity shows proportion of patients with positive DKS, top pairs labeled, legend at the top. **J-K)** Disjoint Killing by Drug Class and Mechanism: Compares DKS (\log_{10} value on y-axis) by general drug classes ($N = 3$ chemo+chemo, 7 chemo+targeted, 5 targeted+ targeted) (**J**) and mechanisms of action ($N = 3$ each MOA) (**K**). Evaluated by two-sided Wilcoxon rank-sum test. Box plots show median, 25th/75th percentiles, and range. **L)** Clone-Level Response in Lung Cancer: Shows post-treatment viability for top effective combinations, one facet per patient. Color intensity indicates clone viability (x-axis) for each drug (y-axis), for the top three patients ranked by highest DKS score per drug.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

We used open-source R versions 4.0 through 4.2 (whatever was the current version at the time) to collect data. To our knowledge, the changes introduced in R between 4.0 and 4.2 have no effect on the results.

Data analysis

The study's scripts to replicate each step of results and plots can be accessed via the following GitHub repository (https://github.com/ruppinnlab/SCPO_submission). We used open-source R versions 4.0 through 4.2 to generate the figures. Wherever required, commercially available Adobe Illustrator was used to create the figure grids.

We use the following libraries for the current versions of analysis and Figures: RColorBrewer 1.1-3, ggplot2 3.4.3, ggpunr 0.6.0, gridExtra 2.3, grid and lattice (tied to R 4.2), data.table 1.14.8, tictoc 1.2, statar 0.7.6, parallel (tied to R 4.2), glmnet 4.1-8, caret 6.0-94, randomForest 4.7-1.1, tidyr 1.3.0, ggrepel 0.9.3, pROC 1.18.4, scales 1.2.1, dplyr 1.1.2, viridis 0.6.4, reshape2 1.4.4, Seurat 4.3.0.1, corrplot 0.92, devtools 2.4.5, and readxl 1.4.3. These libraries are integral for managing, analyzing, and visualizing single-cell omics data to predict patient treatment responses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The entire collection of the processed datasets used in this manuscript, including pre-clinical models of cancer cell lines and PDCs, can be accessed via the following Zenodo repository (<https://zenodo.org/record/7860559>). We collected the bulk-expression and drug response profiles generated in cancer cell lines curated from <https://depmap.org/portal/download/> (version 20Q1). The sc-expression of 205 cancer cell lines was generated in a previous study³⁴ and was downloaded from: https://singlecell.broadinstitute.org/single_cell/study/SCP542/pan-cancer-cell-line-heterogeneity#study-download. The sc-expression profiles of multiple myeloma patients were downloaded from the original study Supplementary Table 2 (https://static-content.springer.com/esm/art%3A10.1038%2Fs41591-021-01232-w/MediaObjects/41591_2021_1232_MOESM3_ESM.xlsx), of breast cancer were downloaded from GEO (GSE158724), and, of NSCLC patients were provided by the original study authors (Citation 41 from main text).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

The reported sex was used in the study, but no analysis comparing males to females was done due to small numbers of samples available. We also considered sex as a variable in our regression wherever available during modeling drug response.

Reported sex information for the three trials used can be found here: 1. NCT04065789 (https://static-content.springer.com/esm/art%3A10.1038%2Fs41591-021-01332-w/MediaObjects/41591_2021_1232_MOESM1_ESM.pdf), NCT02712723 (https://static-content.springer.com/esm/art%3A10.1038%2Fs43018-021-00215-7/MediaObjects/43018_2021_215_MOESM2_ESM.pdf), NCT03433469 (<https://clinicaltrials.gov/ct2/show/study/NCT03433469?term=NCT03433469&draw=2&rank=1>), NCT03088930 (<https://clinicaltrials.gov/ct2/show/NCT03088930>)

Reporting on race, ethnicity, or other socially relevant groupings

We do not have access to patient's race, ethnicity or socially relevant grouping for the multiple myeloma or breast cancer data sets. For the lung cancer data set, we could get access, but the sample size is too small to do any meaningful analyses. For the cell line data, we did not do analysis by sex because the cell lines are highly transformed, which is a noted exception in the SAGER guidelines.

Population characteristics

Population characteristics using the following link. 1. NCT04065789 (https://static-content.springer.com/esm/art%3A10.1038%2Fs41591-021-01332-w/MediaObjects/41591_2021_1232_MOESM1_ESM.pdf), NCT02712723 (https://static-content.springer.com/esm/art%3A10.1038%2Fs43018-021-00215-7/MediaObjects/43018_2021_215_MOESM2_ESM.pdf), NCT03433469 (<https://clinicaltrials.gov/ct2/show/study/NCT03433469?term=NCT03433469&draw=2&rank=1>), NCT03088930 (<https://clinicaltrials.gov/ct2/show/NCT03088930>)

Recruitment

We did not contribute in the design of the trials. The outliers and exclusion criteria can be accessed using the above link (Population Characteristics).

Ethics oversight

NA.
The clinical trials data used in the study was downloaded from published studies: Cohen et al. 2020; Griffiths et al. 2021, December 9th, 2021.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No prior tool was used to determine sample size. We downloaded and used all the sc-expression profiles to us for both patients and cell lines. For cell lines, we conclude the the number of cell line is enough for at least 44 drugs, where we were able to build a predictive model. For the rest of the drugs, we concluded that we do not have enough sample size to build a predictive models.

Data exclusions

No data was excluded either during training or testing.

Replication

All analysis in the study are reproducible and all except two are deterministic. We provided a detailed and single-file execution to help users

Replication

replicate our studies and have tested the reproducibility with other co-authors.

We note that clustering of sc-RNA-seq is non-deterministic and can see slight changes every time it is repeated due to the non-deterministic methods used. However, to ensure that users can replicate our findings, we provide the precise seeds used and also repeated this analysis 1000 times to ensure robustness of our findings.

Randomization

No randomization was performed for primary analysis of the clinical trial data in the usual meaning of 'randomization in clinical trials'. However, for our model building control, we built random models using either shuffled labels, randomized features in the regression model, or a non-predictive model of another drug in the screen for 1000 times and computed the number of times that the stratification power denoted by AUC is higher than our original model. The comparison to other methods also used random sampling and replicates as described in Methods (after the revision requested by the Editor on this point specifically).

Blinding

No blinding was performed. The clinical data analysis is retrospective and thus no blinding was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-----|---|
| n/a | Involved in the study
<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/> Plants |
|-----|---|

Methods

- | | |
|-----|---|
| n/a | Involved in the study
<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/> MRI-based neuroimaging |
|-----|---|

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.